DOCUMENT RESUME

ED 103 467 TM 004 312

TITLE	National Assessment of Educational Progress. General Information Yearbook.
INSTITUTION	Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.
SPONS AGENCY	Carnegie Corp. of New York, N.Y.; Ford Foundation, New York, N.Y.; National Center for Educational Statistics (DHEW/OE), Washington, D.C.
REPORT NO	03-04-GIY
PUB DATE	Dec 74
NOTE	58p.
AVAILABLE FROM	Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402
EDRS PRICE	MF-$0.76 HC-$3.32 PLUS POSTAGE
DESCRIPTORS	*Academic Achievement; *Data Analysis; Data Collection; Data Processing; Educational Accountability; *Educational Assessment; *Educational Objectives; Elementary Secondary Education; Information Dissemination; Mathematics; *Measurement Goals; Measurement Techniques; Models; Music; National Surveys; Sampling; Sciences; Social Studies

ABSTRACT

This second General Information Yearbook published by the National Assessment of Educational Progress (NAEP) is designed to assist the reader in understanding Assessment findings. This report relates especially to the learning areas assessed during the third and fourth years of the project: social studies, music, mathematics and science. It presents an overview of National Assessment, describing its purpose, history, and methodology, and then goes on to devote a chapter to each of these major steps in the intricate research/dissemination process: overview; sampling; objectives and exercise development; preparation of materials and administration of the assessment; categorizing responses and preparing the data files; data processing; data analysis; reporting and disseminating the findings; and adaptation and utilization of the assessment model. Appendices contain the definitions of NAEP reporting populations, estimated population distributions on NAEP reporting groups, a list of the NAEP policy committee and advisory committee. (Author/RC)

BEST COPY AVAILABLE

TM

ED103467

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

General Information Yearbook

TM 004 312

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS
A Project of the Education Commission of the States

David Hall, Governor of Oklahoma, Chairman, Education Commission of the States
Wendell H. Pierce, Executive Director, Education Commission of the States
J. Stanley Ahmann, Director, National Assessment

Assessment Reports

Science, 1969–70		
1	National Results, July 1970	$1.75
4	Results by sex, region and size of community, April 1971	1.00
7	Results by color, parental education, size and type of community; balanced results, May 1973	2.60
Citizenship, 1969–70		
2	National Results, November 1970	1.70
6	Results by sex, region and size of community, July 1971	1.25
9	Results by color, parental education, size and type of community, May 1972	2.85
Writing, 1969–70		
3	National Results, November 1970	1.50
5	Results by sex, region and size of community, April 1971	1.25
8	Writing Mechanics, February 1972	1.75
10	Selected Essays and Letters, November 1972	8.00
11	Results by color, parental education, size and type of community, sex and region for objectively scored exercises, May 1973	1.25
02-GIY	General Information Yearbook—A description of National Assessment's methodology, with special attention given to Reading and Literature, May 1972	1.50
Reading, 1970–71		
02-R-01	Understanding Words and Word Relationships, April 1973	1.25
02-R-02	Graphic Materials, June 1973	2.60
02-R-03	Written Directions, May 1973	2.10
02-R-04	Reference Materials, July 1973	2.15
02-R-05	Gleaning Significant Facts From Passages, May 1973	2.60
02-R-06	Main Ideas and Organization, July 1973	2.10
02-R-07	Drawing Inferences, August 1973	2.60
02-R-08	Critical Reading, May 1973	2.35
02-R-09	Reading Rate and Comprehension, December 1972	2.85
02-R-20	Released Exercises, July 1973	3.95
02-R-30	Recipes, Wrappers, Reasoning and Rate: A Digest of the First Reading Assessment, April 1974	1.00
02-R-00	Summary Data, July 1974	.70
Literature, 1970–71		
02-L-01	Understanding Imaginative Language, March 1973	2.65
02-L-02	Responding to Literature, April 1973	2.85
02-L-03	Recognizing Literary Works and Characters, April 1973	2.60
02-L-04	A Survey of Reading Habits, May 1973	2.85
02-L-20	Released Exercises, April 1973	3.45
02-L-00	Summary Data, June 1973	1.25
Social Studies, 1971–72		
03-SS-01	Political Knowledge and Attitudes, December 1973	1.05
03-SS-02	Contemporary Social Issues, July 1974	.65
03-SS-00	The First Social Studies Assessment: An Overview, June 1974	.90
Music, 1971–72		
03-MU-01	The First National Assessment of Musical Performance, February 1974	.55
03-MU-02	A Perspective on the First Music Assessment, April 1974	.45
03-MU-03	An Assessment of Attitudes Toward Music, September 1974	.85
03-MU-00	The First Music Assessment: An Overview, August 1974	.60
03/04-GIY	General Information Yearbook—A description of National Assessment's methodology, with special attention given to Music, Social Studies, Science and Mathematics, December 1974	1.20

ED103467

BEST COPY AVAILABLE

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

GENERAL INFORMATION YEARBOOK

Report No. 03/04-GIY

December 1974

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

J. Stanley Ahmann
Director

George H. Johnson
Associate Director

This publication was prepared and produced pursuant to agreements with the National Center for Educational Statistics of the U.S. Office of Education with additional funds from the Carnegie Corporation of New York and the Ford Foundation's Fund for the Advancement of Education. The statements and views expressed herein do not necessarily reflect the position and policy of the U.S. Office of Education or other grantors but are solely the responsibility of the National Assessment of Educational Progress, a project of the Education Commission of the States.

TABLE OF CONTENTS

LIST OF EXHIBITS v

PREFACE vii

CHAPTER 1 Overview of National Assessment 1
History and Purpose 1
Goals of the Assessment 2
Methodology 2
The Assessment Cycle 4

CHAPTER 2 Sampling 5
Design Specifications 6
In-School Sample Design 6
Out-of-School Sample Design 7
Sample Sizes 8
Response Experience 8

CHAPTER 3 Objectives and Exercise Development 13
Development and Review of Objectives 13
Exercise Preparation and Local Tryouts 13
Exercise Review and Revision 14
Field Testing, Exercise Scoring and Review of the Results 14
Final Exercise Review and Selection 15

CHAPTER 4 Preparation of Materials and Administration of the Assessment 17
Preparation of Materials 17
Administration of the Assessment 18

CHAPTER 5 Categorizing Responses and Preparing the Data Files 21
Receipt Control 21
Ongoing Hand Scoring 21
Programming and Machine Processing 22
Special Scoring 23
Special Projects 23

CHAPTER 6 Data Processing 25
Computer Hardware and Languages 25
Preparation for an Assessment Year 25
Organizing the Data 26
Availability of Data 27

CHAPTER 7 Data Analysis 29
Computation of P-Values 29
ΔP Values 30
Reliable Differences 30
Summary Results 31
Estimation of Change 32
Limitations of the Data 32

CHAPTER 8 Reporting and Disseminating the Findings 33

CHAPTER 9 Adaptation and Utilization of the Assessment Model 35
Consultation 35
Technical Assistance 36
Workshops and Seminars 36

APPENDIX A Definitions of National Assessment Reporting Populations 39
National Assessment Groups 40

APPENDIX B Estimated Population Distributions of National Assessment Reporting Groups . . 43

APPENDIX C National Assessment Policy Committee 53

APPENDIX D National Assessment Analysis Advisory Committee 55

LIST OF EXHIBITS

EXHIBIT 2-1. National Assessment Reporting Groups 5

EXHIBIT 2-2. National Assessment Sampling Strata 7

EXHIBIT 2-3. Year 03 (Music and Social Studies) Observed Sample Size 9

EXHIBIT 2-4. Year 04 (Science and Mathematics) Observed Sample Size 9

EXHIBIT 2-5. Average Number of Responses by Group and Type of Administration at Each Age Level—Year 03 (Music and Social Studies) 10

EXHIBIT 2-6. Average Number of Responses by Group and Type of Administration at Each Age Level—Year 04 (Science and Mathematics) 11

EXHIBIT 7-1. How Differences from the National Percentage of Success are Reported: Sample Graph 31

EXHIBIT A-1. National Assessment Populations, Survey Periods and Age Ranges 39

EXHIBIT A-2. Definitions of National Assessment Regional Subpopulations 40

EXHIBIT A-3. Occupational Categories 41

EXHIBIT A-4. National Assessment Size and Type of Community (STOC) Reporting Categories 42

EXHIBIT A-5. Distribution of Respondents by Size and Type of Community (STOC) and Age Level 42

EXHIBIT B-1. Estimated Population Distributions of National Assessment Reporting Groups (9-Year-Olds) 44

EXHIBIT B-2. Estimated Population Distributions of National Assessment Reporting Groups (13-Year-Olds) 46

EXHIBIT B-3. Estimated Population Distributions of National Assessment Reporting Groups (All 17-Year-Olds) 48

EXHIBIT B-4. Estimated Population Distributions of National Assessment Reporting Groups (Young Adults) 50

PREFACE

This is the second General Information Yearbook published by the National Assessment of Educational Progress (NAEP). It is designed to assist the reader in understanding Assessment findings. Report 03/04-GIY relates especially to the learning areas assessed during the third and fourth years of the project: social studies, music, mathematics and science. It presents an overview of National Assessment, describing its purpose, history and methodology, and then goes on to devote a chapter to each of the major steps in the intricate research/dissemination process.

The chapters of the General Information Yearbook are organized to tell how the assessment is carried out and subsequently how its data are reported and used. The appendixes provide a more detailed description of procedures, groups, reporting categories and persons involved. For more technical descriptions of National Assessment procedures, see our monographs entitled *The National Assessment Approach to Exercise Development* (1970) and *The National Assessment Approach to Sampling* (1974).

Reporting and Numbering System

For the first assessment year, 1969—70, results were reported primarily in statistical volumes numbered chronologically. National results for science, for example, are found in Report 1, citizenship in Report 2 and writing in Report 3. Results are reported by groups in subsequent reports.

Reading and literature, the learning areas assessed in Year 02, 1970—71, are reported by themes or topics. At that time, a new numbering system was instituted which indicates the assessment year, the learning area and the type of data that will be found inside. The new report number consists of three parts. The first part is a two-digit number referring to the assessment year—for example, 02 for the second assessment year. The second part is an initial letter or letters for the learning areas as follows: A for art, COD for career and occupational development, C for citizenship, L for literature, MA for mathematics, MU for music, R for reading, S for science, SS for social studies and W for writing.

The third part is a two-digit number indicating the content of the volume as follows:

00 = a summary or overview of a learning area

01—19 = results for a theme or topic within a learning area

20 = released exercise or statistical report

21—29 = research or special reports as applicable

30 = a digest or condensation of findings

For example, the Reading Summary is Report 02-R-00, the Social Studies Overview is 03-SS-00, the statistical report containing the exercises with results for mathematics will be numbered 04-MA-20.

How to Secure Reports

The front inside cover of this report lists all of the reports issued by National Assessment to date. Although limited copies are kept at the Denver office for consultants, committee members and media representatives, NAEP reports should be ordered directly from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. When ordering, check must be enclosed as well as full title, e.g., National Assessment of Educational Progress, Report No. 3, Writing — National Results.

CHAPTER 1

OVERVIEW OF NATIONAL ASSESSMENT

History and Purpose

When the United States Office of Education was founded in 1867, one charge set before its commissioner was to determine the nation's progress in education. That century-old charge is only now being answered by the National Assessment of Educational Progress (NAEP), a project of the Education Commission of the States.

By the early 1960s, the average annual expenditure of public funds for the formal education of young Americans was $30 billion.[1] Yet criticism of the educational system abounded. Defenders of the educational establishment found it increasingly difficult to provide evidence that the schools were satisfactorily meeting the educational needs of a modern, technological society. The only readily available measures of educational quality resulting from this public investment of funds were *input* measures such as teacher-student ratios and per-pupil expenditures. The tenuous assumption was made that the quality of educational *outcomes*—what students do or do not know and can or cannot do—was directly related to the quality of the inputs to the educational system. There has been no conclusive empirical evidence to support this assumption. The typical standardized achievement tests administered by schools or states provided scores whereby one student could be compared with other students. Such information was useful in categorizing students; however, it provided little information about what students were or were not actually learning. There had been no direct assessment of educational outcomes.

This insufficiency of information became the concern of Dr. Francis Keppel, United States commissioner of education (1962—65). He initiated a series of conferences to explore ways to provide the necessary information. In 1964, as a result of these conferences, John W. Gardner, president of the Carnegie Corporation, asked a distinguished group of educators and other concerned persons to form the Exploratory Committee on Assessing the Progress of Education (ECAPE). This committee, chaired by Dr. Ralph W. Tyler who had been involved since the earliest conferences, was to examine the possibility of conducting an assessment of educational attainments on a national basis.

After much study, ECAPE decided that it was feasible to inaugurate an assessment project to fill the information gap regarding the quality of educational outcomes by a periodic assessment of the knowledges, understandings, skills and attitudes in 10 learning areas[2] at four age levels (9, 13, 17 and young adult—ages 26—35). This project—named the Committee on Assessing the Progress of Education (CAPE)—began its charge under the auspices of the Carnegie Corporation by assessing the learning areas of science, citizenship and writing in the spring of 1969. Later that same year, the project came under the auspices of the Education Commission of the States. Funding and monitoring were transferred to the United States Office of Educa-

[1] Based on United States Census Bureau data. Actual amounts expended were: 1960, $24.7 billion; 1962, $29.4 billion; 1964, $35.9 billion.

[2] Art, career and occupational development, citizenship, literature, mathematics, music, reading, science, social studies and writing.

tion, and the project was renamed the National Assessment of Educational Progress (NAEP).

Goals of the Assessment

National Assessment provides information to educational decisionmakers and practitioners that can be used to identify educational problem areas, to establish educational priorities and to determine the national progress in education. To do so, NAEP must remain flexible enough to accommodate possible extensions, refinements and modifications. The following goals have been established for the project by the National Assessment Policy Committee, the Analysis Advisory Committee[3] and the NAEP staff.

Goal I: To measure change in the educational attainments of young Americans.

Goal II: To make available on a continuing basis comprehensive data on the educational attainments of young Americans.

Goal III: To utilize the capabilities of National Assessment to conduct special interest "probes" into selected areas of educational attainment.

Goal IV: To provide data, analyses and reports understandable to, interpretable by and responsive to the needs of a variety of audiences.

Goal V: To encourage and facilitate interpretive studies of NAEP data, thereby generating implications useful to educational practitioners and decisionmakers.

Goal VI: To facilitate the use of NAEP technology at state and local levels when appropriate.

Goal VII: To continue to develop, test and refine the technologies necessary for gathering and analyzing NAEP achievement data.

Goal VIII: To conduct an ongoing program of research and operational studies necessary for the resolution of problems and refinement of the NAEP model. (Implicit in this goal is the conduct of research to support previously mentioned goals.)

Methodology

To measure the nation's educational progress, National Assessment estimates the percentage of respondents (at four age levels) who are able to answer a question acceptably or perform a task. Each question or task (called an exercise) reflects a previously defined educational goal or objective. The exercises are administered to scientifically selected samples (which take into account size of community and socioeconomic status and include respondents from all 50 states). Students are sampled at three age levels that represent educational milestones attained by most students: age 9, when most students have been exposed to the basic program of primary education; age 13, when most students have finished their elementary school education; and age 17, when most students are still in school and completing their secondary education. To accurately reflect the skills, knowledge and attitudes of the 17-year-olds, National Assessment also samples 17-year-olds not enrolled in school. Young adults (ages 26 to 35) are assessed to determine the skills, knowledge and attitudes of those who have completed their formal education and have probably been away from school for a number of years. The samples are designed so that sound inferences can be made about the populations from which the samples were selected.

[3] Names of the Policy Committee and the Analysis Advisory Committee members are given in Appendix C and Appendix D, respectively.

NAEP does not develop or use scores for individual respondents. Rather, it determines how the four age levels perform on specific exercises and, within each age level, how groups of individuals (based on demographic and sociological variables) perform. Thus, it is not necessary for each respondent to take every exercise. The exercises are divided into booklets, and each in-school respondent takes only one booklet. Since the samples for the different booklets are statistically equivalent, group comparisons can be made across booklets. This allows National Assessment to assess performance on far more exercises in a learning area than would be possible in the usual one-hour testing situation and provides broader coverage of the Assessment objectives for each learning area.

While multiple-choice exercises predominate, many open-ended exercises requiring anywhere from a few words to a long essay as an answer are included in each assessment. Exercise writers are instructed to use the exercise format that provides the best and most direct measure of the objective being assessed. They are encouraged to develop exercises that employ the use of pictures, tapes, films or practical, everyday items as stimuli. Individual interviews, the manipulation of apparatus to solve a problem and observations of the respondents' problem-solving techniques are used to supplement the usual paper-and-pencil tasks. For example, in music respondents were asked to sing a song or perform on an instrument; in science respondents were asked to conduct a small experiment; in mathematics respondents were asked to make change from a change drawer; in social studies respondents were asked to interpret an election ballot.

NAEP regards positive attitudes toward or opinions about the various learning areas as important educational attainments. Therefore, affective exercises and attitude survey questions are also included in each assessment.

Since individuals are not ranked according to their performance on the assessment materials, National Assessment does not emphasize the use of exercises with high discrimination power. The aim of the project is to describe attainment; this is best accomplished if the exercises used cover the entire spectrum of difficulty, from very easy tasks to the most difficult.

Assessment exercises are administered either to individuals or to small groups (never larger than 12) by specially trained personnel.[4] Exercises specifically designed for individual administration include those having unusual stimuli or requiring something other than a written response. Some exercises at age 9 are administered individually so the youngest respondents can have the opportunity to express themselves verbally instead of by writing. In group administrations, all instructions and the exercises themselves are presented to the respondents on paced tape recordings prepared by professional tape script readers. This assures complete and uniform presentation of instructions and gives those who have a reading problem a chance to hear the exercise as they are reading it.

To report the nation's educational progress, the project releases approximately one half of the exercises administered in a learning area. This allows the public to see the exercises and to evaluate the accompanying data in light of each exercise and its format. The other half of the exercises are kept confidential and are used to assess performance changes over time on specific exercises in the learning area.

How does NAEP differ from standardized achievement tests? Standardized achievement tests are norm referenced; National Assessment is content or objective referenced. With a standardized achievement test, each respondent takes every exercise, receives a score for his performance and is ranked on the basis of that score with respect to a reference

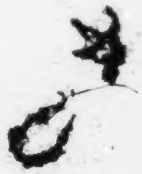
[4] Exercises are administered to all out-of-school respondents in a one-to-one situation. About one fourth of the in-school respondents are assessed in a one-to-one situation. However, this proportion varies from age level to age level and from assessment year to assessment year.

group. With National Assessment no respondent takes all the exercises used to assess a learning area, no respondent receives a score and emphasis is placed on the performance of groups of respondents on specific exercises. A standardized achievement test is usually mass-administered; NAEP exercises are administered to small groups of no more than 12 and also contain individually administered sections. Standardized achievement test items are usually limited to a multiple-choice format; National Assessment employs a wide variety of exercise formats. Standardized tests usually focus on the cognitive domain; NAEP includes exercises relating to the affective domain as well. A respondent is required to read the items himself when taking most standardized tests; Assessment exercises are read to respondents by a paced tape or the exercise administrator in an interview situation except during the reading assessment. The items on a standardized achievement test are rarely, if ever, made public; the Assessment releases half of the exercises used in an assessment to accompany the data.

The Assessment Cycle

Much of the work of National Assessment is concerned with the year by year implementation of an overall design and schedule which was developed at the inception of the project. Thus, many of the activities are baseline activities, repeated for each learning area and for each cycle in accordance with an established design and employing basically the same procedures and technology.

A single cycle of a learning area assessment, from objectives development or redevelopment to completion of the basic technical reporting of the data, requires approximately 6 years. Development takes 3½ years; 1½ years are spent preparing for and performing data collection, and 1 year is required for preliminary analysis and basic reporting.



CHAPTER 2

SAMPLING[1]

National Assessment's (NAEP) principal objective is to assess the changes over time in educational achievement of four age groups in the United States. In addition to the national results for each age group, results are reported within each age level for certain major subpopulations, such as geographic region and sex. Current reporting populations and subpopulations are shown in Exhibit 2-1; the definitions for each are presented in Appendix A.

EXHIBIT 2-1. National Assessment Reporting Groups

Classification	Subgroup
Age level	9-year-olds 13-year-olds 17-year-olds Young adults (26-35 years)
Sex	Male Female
Geographic region	Northeast Southeast Central West
Level of parental education	No high school Some high school Graduated high school Post high school
Size and type of community (STOC)	Low metropolitan High metropolitan Extreme rural Main big city Urban fringe Medium city Small places
Color	White Black

There are two ways to meet this objective: conduct a census of all members of the four age populations or utilize probability samples of the four populations. A complete census has the advantages of providing great precision and of allowing very detailed reporting. However, the four populations are very large (about 37 million people altogether); finding and assessing that many people would involve enormous expenditures of time and resources for data collection, scoring and analysis. Furthermore, it would be extremely difficult to adequately train and supervise the large field staff necessary to assess all members of the population.

Since National Assessment only collects data about a limited number of subpopulations and not individual students, schools, school districts or even states, relatively small probability samples allow estimation of educational performance in these subpopulations with the required degree of precision. Further, by limiting the scale of data collection activities, it is possible to use a small, highly trained professional field staff to collect data of higher quality than would be possible with a census. Quantity of data is traded for much higher quality at reasonable cost.

[1] This chapter provides a brief overview of National Assessment sample designs for assessment Years 02—04. For more detailed documentation, see J.R. Chromy et al., *The National Assessment Approach to Sampling* (Denver, Colorado: National Assessment of Educational Progress, 1974). For an introductory treatment aimed at state and local assessment problems, see R.M. Jaeger, *A Primer on Sampling for Statewide Assessment* (Princeton, N.J.: Educational Testing Service, Center for Statewide Educational Assessment, 1973).

BEST COPY AVAILABLE

The development of sample designs was based upon a long series of interactions between what the Assessment ought optimally to accomplish and what it could accomplish, given the available finite resources. Policy, data collection, analysis, reporting, cost and time considerations determined what types of sample designs were possible. The sample designs, in turn, put major constraints on the options available for data collection, analysis and reporting. For example, the limits placed on the number of administrations in a school and the size of group administrations (12 students) are a compromise between cost efficiency and practical feasibility considerations.

Design Specifications

The target populations in the Year 03 and 04 assessments included 9, 13 and 17-year-olds enrolled in public or private school, 17-year-olds who either left school before graduating or graduated early and young adults 26 to 35 years old. Age-eligible persons who were non-English speaking, institutionalized or handicapped (physically, mentally or emotionally) in such a way that they could not respond to the exercises as administered, were excluded.

At each age level the sample for each package of exercises was designed to meet the following specifications:

1. Adequate representation of the subpopulations to allow estimation of the desired proportions with an acceptable level of precision.

2. Representation by at least one sample point for each of the states and the District of Columbia. The design was not, however, to provide for making comparisons among states, school districts, schools, teachers or individual students.

3. Facilitation of field-operating procedures.

NAEP's policy that administration of materials was not to take more than one class period of a student's time, that the demands on school personnel were to be minimized (one package per student) and that the number of students assessed in any one school was to be limited to 12 also placed constraints on the sample design. Among other important sampling considerations were (1) that no more than about one half of the group-administered packages were to be administered in any one school, (2) that out-of-school 17-year-olds and young adults were to be allowed to take up to four packages and (3) that the respondents taking each package were to be a probability subsample of the total sample for the age group.

This design provided for simple, precise estimates of population proportions. Since results were to be reported by individual exercise, these estimates were to be relatively inexpensive to compute.

To meet these specifications, two three-stage, deeply stratified cluster designs with extra sampling of certain strata were developed. One design was for the survey of 9, 13 and 17-year-olds enrolled in school. The out-of-school design was developed to assess young adults aged 26 to 35 and 17-year-olds not enrolled in school.

In-School Sample Design

Stratification

In a stratified design the population is divided into two or more groups, or strata. Samples are then drawn from each stratum rather than from the population as a whole. The two major reasons for stratification are to insure representation of specified subpopulations and to achieve more precise estimates. National Assessment samples were stratified by geographic region, state, size of community and socioeconomic level as shown in Exhibit 2-2. Socioeconomic stratification was performed at the school level in the two largest size of community (SOC) strata and at

the primary sampling unit level in the remaining two size of community strata.

EXHIBIT 2-2. National Assessment Sampling Strata

Classification	Strata
Region	Northeast
	Southeast
	Central
	West
Size of community (SOC)	Counties containing cities with 1970 populations of 180,000 or greater
	Counties in the same Standard Metropolitan Statistical Area (SMSA) as the above counties
	Counties not in the above strata that are part of an SMSA or contain a city with a 1970 population of 25,000 or greater
	All other counties
Socioeconomic level	High–schools in the first two SOC strata
	Low–counties in the last two SOC strata

Multi-Stage Sampling

A multi-stage design involves sampling in successive steps or stages in order to control sampling and data collection costs.

In National Assessment's in-school sample design, the first stages—or primary sampling units (PSUs)—consisted of counties or groups of contiguous counties. A sampling frame of PSUs was constructed with United States Census data on the number of persons by age in each PSU. The PSUs were stratified by region, and within region by state, size of community and, for the two smaller size of community strata, by socioeconomic level. From this stratified list of PSUs, a probability sample of 116 PSUs was independently drawn each year.

At the second stage, a list was made of all public and private schools in each selected PSU. The listing included the estimated number of 9, 13 and 17-year-olds enrolled in each school. A secondary sample of schools was then drawn for each sample PSU. Schools in the two larger size of community strata were stratified by socioeconomic level prior to selecting the secondary sample.

In each selected school, the third-stage sampling units consisted of eligible students enrolled in the school. Every eligible student was listed. A random sample of students was then drawn and randomly assigned one of the assessment packages scheduled for that school.

To ensure adequate sample sizes for each of the various reporting groups listed in Exhibit 2-1, it was necessary to increase the sample size of the population at the lower end of the socioeconomic scale. This "oversampling" was performed at the first stage in the two smaller size of community strata and at the second stage in the two remaining size of community strata.

Out-of-School Sample Design

The out-of-school sample design is similar to the in-school design. As with the in-school design, primary sampling units consisted of counties or groups of contiguous counties stratified by region, state, size of community and socioeconomic level. One hundred four primary sampling units were drawn and used for both Year 03 and Year 04 (one half of them were used in Year 02). Each year, a new secondary sample of small land areas (segments) was drawn in each PSU. All of the housing units in each segment were listed and a sample of the housing units was screened for eligible out-of-school 17-year-olds and adults.

The Supplementary Frame

In order to increase the sample size of 17-year-olds not enrolled in school, a probability subsample of one half of the schools in the 17-year-old assessment was drawn. Each school was asked to provide lists of persons

who had either graduated early or dropped out of school. A subsample of dropouts and early graduates was drawn.

Sample Sizes

The size of a sample required for estimating baseline performance and change over time is related to the type of administration, the minimum change to be detected with a given degree of confidence, the desired power, the sample design, operational procedures, time and cost. In accounting for these, NAEP set the planned sample size for the in-school samples at 12 individuals per group-administration session and the planned national sample sizes at 2,592 for group-administered packages and 2,160 for individually administered packages.

One problem recognized in planning the sample sizes for the in-school assessment was the reduction in sample size due to absenteeism at the time of assessment. Therefore, to assure that the desired sample sizes were achieved, a random sample of 16 students was selected for each group administration. The first 12 students were assigned to the assessment session. The remaining 4, designated as alternates, were used to replace any of the 12 students who were absent at the time scheduled for the package administration. For each individual administration, 2 students were randomly selected with the second serving as an alternate for the first. Special selection procedures were adopted for the in-school sample to accommodate schools enrolling less than the required number of students for one group administration. The allocations to those schools were in addition to the planned national sample sizes given above.

Response Experience

The response experiences for Year 03 and Year 04 are summarized in Exhibits 2-3 through 2-6. The observed sample sizes per package are presented in Exhibits 2-3 and 2-4. Exhibits 2-5 and 2-6 present the average size per package for each of the reporting groups. The planned sample sizes were 2,592 respondents per package for in-school group administrations, 2,160 per package for in-school individualized administrations and 2,000 respondents for each package in the out-of-school sample. The exhibits reveal that the observed sample sizes were very close to the projected sizes.

BEST COPY AVAILABLE

EXHIBIT 2-3. Year 03 (Music and Social Studies) Observed Sample Size

Package Number	Age Level			
	9	13	17	Adult
1	2,717	2,661	2,684	2,353
2	2,712	2,665	2,679	2,328
3	2,733	2,600	2,654	2,350
4	2,661	2,664	2,707	2,335
5	2,296*	2,667	2,672	2,367
6	2,321	2,741	2,659	2,334
7	2,341	2,691	2,665	2,356
8		2,286*	2,670	2,337
9		2,297	2,722	
10		2,287	2,498*	
11			2,498	
12			2,497	
Packages administered:	17,781	25,569	31,605	18,760
Respondents:	17,781	25,569	28,812	4,822

Total number of packages administered at all ages: 93,715
Total number of respondents at all ages: 76,984†

**Sample sizes below line are for individually administered packages.*
†Adults and out-of-school 17-year-olds were allowed to take more than one package

---

EXHIBIT 2-4. Year 04 (Science and Mathematics) Observed Sample Size

Package Number	Age Level			
	9	13	17	Adult
1	2,666	2,531	2,651	1,993
2	2,670	2,648	2,713	1,987
3	2,628	2,674	2,756	1,999
4	2,646	2,621	2,721	2,004
5	2,670	2,684	2,751	2,030
6	2,674	2,512	2,725	2,036
7	2,684	2,689	2,672	2,034
8	2,258*	2,575	2,682	2,025
9	2,283	2,573	2,638	
10	2,224	2,269*	2,633	
11		2,254	2,643	
12		2,217	2,507*	
13			2,502	
14			2,498	
Packages administered:	25,403	30,247	37,092	16,108
Respondents:	25,403	30,247	33,577	4,211

Total number of packages administered at all ages: 108,850
Total number of respondents at all ages: 93,438†

**Sample sizes below line are for individually administered packages.*
†Adults and out-of-school 17-year-olds were allowed to take more than one package.

EXHIBIT 2-5. Average Number of Responses by Group and Type of Administration at Each Age Level—Year 03 (Music and Social Studies)

	Age Level						
	9		13		17		Adult
	Group	Individual	Group	Individual	Group	Individual	
National	2,306	2,307	2,672	2,290	2,679	2,498	2,345
Region							
Northeast	674	574	651	558	644	601	519
Southeast	696	593	679	582	702	654	582
Central	668	570	681	583	667	622	632
West	669	570	661	567	665	621	612
Sex							
Male	1,338	1,140	1,337	1,146	1,296	1,209	1,060
Female	1,368	1,167	1,335	1,144	1,383	1,289	1,285
Color							
White	2,106	1,796	2,138	1,832	2,177	2,030	1,968
Black	414	353	352	301	335	313	240
Other	186	158	183	157	167	155	137
Parental education							
No high school	157	134	196	168	266	248	705
Some high school	144	122	261	223	373	348	394
Graduated high school	647	552	825	707	816	761	662
Post high school	896	764	1,025	878	1,074	1,002	499
Unknown	862	735	365	313	149	139	84
Size and type of community (STOC)							
Low metro	277	236	264	227	312	291	233
Extreme rural	270	231	255	219	310	289	238
Small place	877	748	899	771	886	826	746
Medium city	454	387	443	379	438	408	321
Main big city	283	242	299	256	240	224	188
Urban fringe	271	231	225	193	144	135	384
High metro	272	232	287	246	349	325	234

EXHIBIT 2-6. Average Number of Responses by Group and Type of Administration at Each Age Level—Year 04 (Science and Mathematics)

	Age Level						
	9		13		17		Adult
	Group	Individual	Group	Individual	Group	Individual	
National	2,663	2,255	2,612	2,247	2,690	2,502	2,014
Region							
Northeast	656	561	651	565	653	620	477
Southeast	669	565	667	565	685	641	508
Central	672	566	649	562	679	618	516
West	665	563	645	555	673	623	512
Sex							
Male	1,328	1,123	1,294	1,098	1,280	1,206	925
Female	1,335	1,132	1,318	1,148	1,409	1,296	1,088
Color							
White	1,997	1,689	1,977	1,683	2,110	1,942	1,688
Black	466	403	436	391	416	413	200
Other	199	163	199	173	163	147	126
Parental education							
No high school	146	114	185	185	248	223	562
Some high school	125	120	232	230	353	335	325
Graduated high school	564	566	792	760	816	782	583
Post high school	787	819	994	837	1,108	1,025	481
Unknown	1,041	637	409	234	164	137	62
Size and type of community (STOC)							
Low metro	265	229	263	222	274	280	200
Extreme rural	266	227	264	221	270	238	197
Small place	891	743	869	740	966	876	847
Medium city	372	326	365	308	380	357	232
Main big city	285	247	277	251	214	204	140
Urban fringe	317	261	314	275	321	300	199
High metro	267	222	260	229	265	247	199

CHAPTER 3

OBJECTIVES AND EXERCISE DEVELOPMENT

The development of objectives and exercises occurs in five steps:

1. development and review of objectives;[1]
2. exercise preparation and local tryouts;
3. exercise review and revision;
4. field testing, exercise scoring and review of the results; and
5. final exercise review and selection.

Development and Review of Objectives

Whether objectives were developed for new areas (math, music, social studies) or redeveloped for previously assessed areas (science) the final version had to satisfy a cross-section of subject-matter specialists and educators in each learning area. Acceptance and approval of each set of objectives as an important set of educational goals was also elicited from concerned citizens. Whenever National Assessment (NAEP) has employed consultants for review conferences and approval of materials, attempts have been made to insure representation from all regions of the country, participation of members of minority groups and participation of both males and females.

In addition, the Assessment selected subject-matter specialists and educators representing different specialties within the discipline; different organizations and projects associated with the learning area or discipline; and both elementary and secondary levels of public, private and parochial school people (classroom teachers, curriculum specialists, administrators).

Further selection requirements for concerned citizens included representation from different types of communities, diverse national organizations and different occupations.

Objectives development took approximately one year to accomplish for each learning area. Contractors and consultants conducted a literature search within each area to document curriculum trends, existing sets of educational objectives and general content organizations. Then subject-matter specialists participated in a series of development and review conferences which yielded a first draft of objectives. Further reviews, including a review by knowledgeable, concerned citizens, were conducted. Once the objectives were revised and approved, the final draft was again reviewed and approved by another group of subject-matter experts. The new concensus objectives were adopted by National Assessment.

Exercise Preparation and Local Tryouts

Based on the objectives for each learning area, NAEP developed specifications for the number, character and quality of exercises to be developed. Special emphasis was given to writer qualifications, documentation, difficulty levels and usefulness at more than one age level. Exercise writers had to be subject-matter experts with experience in the education of 9-year-olds, 13-year-olds, 17-year-olds or adults in the particular learning area.

[1] Objectives development and review in specific learning areas are described in the objectives booklets for those areas.

Exercise documentation had to include a rationale; administration, scoring and reporting directions; scoring keys or scoring categories; sample responses and all special stimuli (e.g., music and the related permissions for use).

Exercises had to be developed with difficulty levels ranging from an expected response rate of over 75% correct responses to an expected rate of less than 25% correct responses. This requirement was made because National Assessment must be able to describe a broad range of educational attainments achieved by the groups of people in the target populations.

Development of exercises that could be administered at two or more age levels was encouraged because of the useful comparisons that could be made between ages.

As part of the exercise preparation process, exercise developers conducted small-scale local tryouts for each exercise they submitted to the project. These tryouts served two functions:

1. They furnished information about exercise clarity and administrative feasibility.

2. They provided sample responses for open-ended exercises which the contractor used to develop initial scoring guides.

Exercise Review and Revision

Newly developed exercises were reviewed by National Assessment staff and the Exercise Development Advisory Group (EDAG), which has a rotating membership comprised of five educational measurement specialists. Subject-matter specialists critiqued each exercise with respect to whether it was a direct, clear measure of an objective for the learning area, whether it was relevant to current educational practices and opinions and whether it contained accurate scoring and reporting guides.

Groups of informed and concerned citizens reviewed each exercise with respect to whether it was an appropriate measure of an objective for a given age, whether it might be offensive to any particular group of people and whether it was a relevant exercise in terms of what they considered to be valuable learning experiences. Criticisms and suggestions were transmitted to the exercise development contractor, who then reviewed all the data, suggestions and criticisms and revised the exercises accordingly.

Field Testing, Exercise Scoring and Review of the Results

Following review and revision, exercises were given a national tryout. Tryout respondents were selected to give representation to the reporting categories: region of the country, size and type of community, socioeconomic levels, color and sex. The tryouts served three purposes:

1. to check administrative feasibilities,

2. to provide data for improvement of scoring guides and

3. to provide data necessary for evaluating exercises in order to select those that would be used in the assessment.

Responses were scored separately for each age. Multiple-choice exercise responses were tallied directly by scoring clerks. Short-answer written responses were categorized according to scoring guide categories, and the number of responses falling in each category was tallied. Recording response data for exercises that involved longer open-ended responses, performance or production was a more complex process. Subject-matter specialists scored these exercises using the following general pattern:

1. reviewing all responses to a given exercise,

2. outlining response categories that were

both relevant to the exercise objective and inclusive of the range of responses,

3. tallying responses occurring in each tentative category,

4. reviewing and revising the categories and

5. tallying responses for the final categories.

Subject-matter specialists carefully reviewed all response data from all exercises. Each scoring guide or exercise was checked for subject-matter accuracy, appropriateness for the age level and biases with respect to any particular group of people. These exercises were also reviewed by NAEP staff members and EDAG. They considered each exercise in light of potential administration, scoring and reporting difficulties, as well as general measurement characteristics. These considerations, taken together with the subject-matter specialists' recommendations, determined the exercises suitable for use in the assessment.

Final Exercise Review and Selection

The pool of suitable exercises had to be narrowed to the number that could actually be used in the assessment. Therefore, the exercises were again reviewed by subject-matter specialists and selected on the basis of exercise quality, importance, difficulty and coverage of desired objectives. This final set of exercises was then forwarded to the United States Office of Education for review and approval.

BEST COPY AVAILABLE

CHAPTER 4

PREPARATION OF MATERIALS AND ADMINISTRATION OF THE ASSESSMENT

Before the data collection for an assessment year begins, materials are prepared for use in the field. Exercises which have been selected for inclusion in the assessment are grouped into booklets—called "packages"—and printed. Tape recordings of the exercises are produced to accompany the packages. Exercise administrators are trained, and materials are shipped to them. All of these activities are planned, coordinated and monitored by National Assessment (NAEP) staff.

In preparation for the music and social studies assessments (1971—72) and the science and math assessments (1972—73), the Research Triangle Institute (RTI) of Raleigh, North Carolina, assembled the exercise packages and produced the tape recordings. The Measurement Research Center (MRC) of Iowa City, Iowa, printed the packages so that the responses, which were marked on the package pages, could be read and recorded by machines. RTI, with a subcontract to MRC for the Central and Western regions, conducted the field administration, and MRC ultimately conducted the scoring and machine processing of the exercise booklets.

Preparation of Materials

Since individuals of different ages were to receive somewhat different sets of exercises, packaging was done separately for each age level in the assessment. Packages were compiled according to the following criteria:

1. Each package contained 35 minutes of exercise time. An additional 15 minutes of administration time was allowed for introductory statements, instructions and completion of personal background questions. For in-school administrations no more than about 50 minutes of a student's time was required for participation in an assessment since each student was allowed to take only one package of materials. Therefore, in most schools a package administration could be completed within the time boundaries of a class period.

2. Each package contained some exercises from both learning areas being assessed that year.

3. Each package contained exercises from the three levels of difficulty (easy, medium, difficult).

4. Exercises could not appear in the same package if they used the same or similar stimulus materials or if one exercise provided a clue to the answer of another.

5. Each package contained exercises covering several topics in each learning area.

Some exercises were designed to be administered to one individual at a time in an interview situation. Most exercises, however, were designed to be completed by respondents on their own as they read the questions and heard them read on tape. The respondents either marked an oval beside a multiple-choice answer or wrote open-ended responses on lines provided in the booklet. Since respondents worked these exercises on their own, up to 12 students could be assessed as a group in one administration. In each assessment year there were three packages of

individual exercises for each age level and from four to nine packages of group exercises.

The package printing procedures added coded and machine-readable identifying information to each exercise page. Quality-control measures insured that all machine-readable information was accurately located on the pages and that paper and ink quality met standards necessary for later machine processing of the exercise booklets.

To assure uniformity in administration, and to help remove biases as well as aid respondents with reading difficulties, tape recordings were produced to accompany each group package. Exercise directions, written stimulus materials, exercise stems and exercise foils were presented both on the tape and in the package. Some exercises in the math assessment were an exception to this rule because it would have been inappropriate to have had certain arithmetic symbols read aloud on tape. The tapes were paced to allow an adequate amount of response time for each exercise. Individual packages generally did not have accompanying tapes, since the administrator read the questions in an interview situation.

Some exercises required the preparation of additional stimulus materials. Several music exercises required that musical stimuli be presented on tapes. Music performance exercises also required that the administrator record responses on a second tape recorder. Individual exercises in other learning areas sometimes involved the use of other materials or equipment, such as pictures, rulers and science apparatus, which were purchased or made especially for use in an assessment.

After all of the materials for the assessment of an age group were prepared, the field staff was instructed about details on how to administer the packages and procedures for contacting schools or locating and sampling individuals for the out-of-school assessment. The assessment year schedule for administration was as follows: 13-year-olds were assessed in the fall (mid-September through mid-December); 9-year-olds were assessed in January and February; and in-school 17-year-olds were assessed in March and April. The assessment of adults and out-of-school 17-year-olds occurred in the spring and early summer months.

Administration of the Assessment

Some aspects of assessment administration procedures varied depending on whether the respondents were in school or out of school. The sampling methods used by the two field staffs were not the same, and the procedures for administration, such as the number of packages administered and the use of tape recordings, also differed.

In-school Administration: 9, 13 and 17-Year-Olds

Before the packages could be administered in the selected schools, cooperation had to be obtained from school personnel, and operational procedures had to be established between them and the administration contractor's field staff. Chief state school officers were informed of all schools selected for the assessment within their respective states. The National Assessment staff director notified superintendents and private school officials that schools from their districts had been selected for participation. The superintendents also received materials from the sampling and administration contractor which identified the selected schools in their districts, described the project and suggested dates for meetings with members of the field staff.

The field staff included 4 regional supervisors and 29 district supervisors. Each district supervisor met with the school officials in his district to explain the purpose of NAEP, describe the operational procedures for completing the assessment and determine a time suitable for assessment in the area. The school officials were also asked to respond to principal's questionnaire items. Questions were asked about the school's enrollment in various grades, the types of communities in

which the students lived, the general occupational levels of the parents in the community and, in the case of the science assessment, the type of science curriculum used by the school.

Each school principal appointed a coordinator who arranged for space to conduct the group and individual administrations and who worked with the district supervisor to arrange a mutually convenient schedule in the school and to ensure that students arrived on time for their scheduled administrations. The coordinator also arranged to provide a listing of each student born during the birth-date range defining National Assessment eligibility. The district supervisor used the listing to make a random selection of the students to be assessed; each student was assigned to receive *one* assessment package.

After the sample was selected in a school, package administrations were done by the district supervisor or by an exercise administrator hired locally and trained by the district supervisor. The exercise administrators had various backgrounds including teaching, substitute teaching and survey research. Assessment time varied between schools depending upon the number of packages assigned to each school; only rarely did administrations within a school take more than three days.

The administrator coded each student's birth date, sex, grade, color and identification number on his or her package. The district supervisor checked all data coded on the packages against the student listing and instructed the school coordinator to save the listing for six months in case it might be needed for data verification. Since names are not associated with NAEP packages, the listing which cross-referenced packages by identification numbers was the only means of verifying lost or questionable data. After the six-month period, the listing was destroyed to protect the anonymity of students who participated in the assessment.

The district supervisor sent the completed packages to the scoring contractor where they were quality checked to verify that correct administrative procedures were being followed by the field staff and that all packages were accounted for. The coded identification information was also checked for accuracy; inconsistencies that could not be reconciled were sent back to the district supervisor to be checked against the student listing.

Each school that participated in the assessment received a thank-you letter and a questionnaire concerning contact procedures, preparation for the assessment, assessment exercises, personal data questions and the school's general involvement during assessment week. This information was analyzed to discover problems in the field procedures so that they could be corrected in future assessments.

Out-of-School Administration: 17-Year-Olds

In order to insure that 17-year-old results would be representative of the entire population of 17-year-olds, efforts were made to locate and assess those 17 year-olds not enrolled in school during the time of the in-school assessment. These 17-year-olds either had withdrawn from school or completed requirements for early graduation.

There were two methods of identifying and locating eligible out-of-school 17-year-olds. First, during their visits to schools for the in-school administrations, the district supervisors obtained lists of out-of-school 17-year-olds. A sample of potential respondents was drawn from these lists and, during the early summer months, the district supervisors attempted to locate and assess these individuals. Second, if the field staff in charge of the adult assessment located any age-eligible 17-year-olds who were not in school, they administered 17-year-old packages to these individuals.

Each year approximately 1,000 out-of-school 17-year-olds were located and assessed. The administration procedures used were the same as those used for the adults.

Out-of-School Administration: Young Adults

Locating young adults ages 26—35 was more difficult than locating students, and a separate, specially trained field staff was assigned to conduct the adult assessment. They used different sampling techniques for selecting respondents and their administrative procedures were different from those of the in-school administration staff.

The field supervisors first identified and visited the areas of the country included in the sample. Next, they compiled a list of all the sample housing units that fell within the area boundaries. The field interviewers then canvassed the sample housing units to find age-eligible adults as well as out-of-school 17-year-olds mentioned previously. The interviewers filled out a screening questionnaire on which they listed all household members and their birth dates. The interview was terminated if the screening questionnaire listed no age-eligible household members. The interviewers made call-backs to households where no one was at home, thereby completing screening questionnaires on more than 99% of the sample housing units.

When an eligible individual was located, the field interviewer explained the amount of involvement that would be required and encouraged participation in the assessment. Handout materials were available to anyone who desired more information about the project. Eligible individuals who refused to participate were later contacted by the field supervisor, and after receiving more information they often agreed to participate. Of those who were contacted in each assessment year approximately 84% of the eligible young adults and 91% of the out-of-school 17-year-olds participated in the assessment.

The one-hour time restriction was not applied to out-of-school administrations. Eligible young adults and out-of-school 17-year-olds were allowed to take as many as four packages of materials. Almost all of the respondents chose to take four packages. They received $5.00 for each package they completed. Package assignment procedures insured that all packages were administered to about the same number of individuals and that each package was administered as the first, second, third or fourth package about the same number of times. Each individual also responded to items on a background questionnaire.

Since the out-of-school assessment took place in the respondents' homes, the administration was always a one-to-one (interviewer and respondent) situation. No packages were administered in a group setting. For some exercises the interviewer asked the questions and wrote down or tape-recorded the individual's responses. Other exercises were "self-administerable" and the respondent read the question while listening to it being read on tape. The tapes were not paced as in the in-school administrations. Instead, the field interviewer started and stopped the tape as necessary to allow the individual a reasonable length of time to respond.

The field interviewers coded the individual's identification information on the packages, and the completed materials were sent to the scoring contractor where quality-control procedures insured that all information was coded correctly and that all of the materials that had been shipped to the field had been returned.

CHAPTER 5

CATEGORIZING RESPONSES AND PREPARING THE DATA FILES

The data that National Assessment (NAEP) collects in the field must be coded into a form suitable for computer processing and analysis. Measurement Research Center (MRC) of Iowa City, Iowa, was the contractor in charge of receiving, categorizing and machine processing the data. Responses to open-ended and performance exercises were categorized and coded into machine-readable form by a special staff of scorers. Responses to multiple-choice exercises were read directly by optical scanning machines. These machines transferred the information to computer data tapes which provided the basis for data analyses performed by the NAEP staff.

Receipt Control

When booklets were printed and shipped to the field staff, each one-week unit of work was packed in a fiberboard box. Once the packages were administered, they were repacked in the same box and returned to MRC for processing and scoring. From the beginning to the end of the MRC processing, that unit of work was kept intact.

Upon receipt from the field, each box was unpacked by the receipt control staff and checked to ensure that:

1. All forms and worksheets were properly and accurately completed.

2. All materials (packages, tapes, work samples, etc.) were properly and uniquely identified.

3. All materials were accounted for.

4. Personal data and documentation for each respondent were complete and consistent.

5. Machine-readable responses were gridded according to instructions and were dark enough to be scorable.

6. Administrator-recorded responses were legible and scorable.

7. Branching exercises were being correctly administered.

MRC corrected errors and inconsistencies which could be solved with the information at hand. Problems for which there was insufficient information were referred to the field staff for solution. In all cases, NAEP staff was notified of the problems and their recommended solutions. After the receipt control checks were completed, each box was repacked and given to the MRC hand-scoring staff.

Ongoing Hand Scoring

In each assessment year, MRC employed a special staff to hand score open-ended exercises. Scorers were responsible for categorizing responses according to predetermined guidelines and coding the category of the response into ovals that could be read by the optical scanning machine.

Hand-scoring tasks ranged from simple tallies to rather high-level judgments. Hand scorers were required to have some academic background and, preferably, teaching experience in the subject they were scoring.

When the initial data (age 13) began to arrive from the field, the scoring staff was assembled for training. A training session took place prior to the scoring of each age level. The first training session was the most complex, since scorers had to become familiar with NAEP, the objectives related to the particular learning area and general scoring and coding procedures.

The MRC director of scoring conducted the training sessions in cooperation with NAEP staff. Once the general orientation was accomplished, training began on specific exercises and their scoring guides. The guides were initially developed by the exercise developer, reviewed and refined by subject-matter and lay consultants and refined further by NAEP staff after tryouts. Acceptable and unacceptable responses were defined through prose descriptions and sample responses. Response categories that were of interest to NAEP staff and consultants were also defined.

Scorers were trained in the use of the guides by scoring sample responses, which were either taken from tryout material or the arriving assessment data. Scorers initially worked as a group and discussed the appropriate categorization of each example response. They then worked individually on another set of responses. Discrepancies were resolved and explained. Once the group felt comfortable using the guides, they started categorizing the actual data. The director of hand scoring checked the scoring for the first 20—24 papers for each exercise for each scorer. Incorrect categorizations were explained and careless errors were pointed out. Subsequently, spot checks were made on the work of each scorer and retraining was conducted as needed.

Based on the final data some scoring guides needed modification. Generally, this involved some redefinition of scoring categories. All responses scored before a scoring guide change were, of course, rescored.

Scorers were approved by NAEP prior to training and could be removed at any time if their work was found unsatisfactory. Quality-assurance procedures were developed both by MRC and NAEP to ensure the consistency of hand scoring. Members of the NAEP staff made several trips to MRC to observe and assist in training sessions and to check that the hand scoring was consistent with the scoring guides furnished by the project.

After all the data at a given age level was scored, MRC prepared a report which included the final form of the scoring guides, examples of responses for scoring guide categories and a summary of the problems encountered and their solutions.

Programming and Machine Processing

Once the open-ended responses were categorized, the packages were repacked and sent on for machine processing. While the exercises were undergoing hand scoring, all of the procedures and programs to be used in the machine processing were being tested on materials developed by MRC and NAEP staff. The optical scanning device was programmed to read the response data on the exercise pages and produce data tapes which could be used by computers.

Once the packages arrived at the MRC processing department, the pages were separated so they could be passed through the scanning device. Each page contained preprinted, machine-readable information such as age, package, exercise and page number. On pages with multiple-choice exercises the respondents had made pencil marks in ovals next to the answers they chose. For other exercises the hand-scoring staff had marked codings in preprinted oval positions on the page. The scanning device, through the use of reflected light, distinguished the position of each pencil-marked oval. The scanner assigned numerical values, specified in computer programs, to the various marks according to their position on the page. The coded information was transferred onto magnetic tapes which later were used to construct the NAEP respondent data files.

MRC's computer checked and edited the contents of the magnetic tapes for package counts, internal exercise consistency and response ranges before they were sent to National Assessment. Tapes were also provided to RTI, the sampling contractor, for calculating weights.[1] After scanning and editing, exercises which required special scoring were isolated. The remaining pages were then sent to a climate-controlled warehouse for permanent storage. It is a policy of NAEP to save all source documents.

Special Scoring

Occasionally, the scoring of an exercise requires special scorer qualifications and training.

Music exercises involving the evaluation of singing and playing an instrument, a mathematics exercise about computer programming and several social studies and science exercises were scored professionally by subject-matter experts. The mathematics exercise was scored by bona fide computer programmers who read only that one exercise. Professional scorers were trained to score all of the musical performance exercises. Judgments were made based on the criteria established in scoring guides. For example, music scorers listened to tape recordings of respondents singing or playing instruments and then scored the renditions for accuracy of pitch and rhythm according to specified criteria.

Special scoring often involved separate scheduling procedures. The ongoing hand scorers left the special exercises blank. These exercises went through the normal machine processing procedures and then were pulled from the packages and held until the time of the special scoring session. These sessions were held after nearly all the regular scoring and processing for the assessment had been completed. After scoring, the exercise pages were rescanned and data tapes were created. The tapes were later combined to construct National Assessment respondent data files.

Special Projects

Each year, NAEP and MRC staffs also carried out various tryouts, feasibility studies and validity studies, which are an integral part of NAEP's development, quality-assurance and feedback program. These special studies varied depending on the requirements of the learning area being assessed.

---

[1] For a discussion of weighting procedures, see Chapter 7, Computation of P-Values.

CHAPTER 6

DATA PROCESSING

The data gathered by Research Triangle Institute (RTI) and Measurement Research Center (MRC) were combined, organized and analyzed by National Assessment (NAEP). Specifications for analyses to be performed were supplied by NAEP's Research and Analysis Department staff. NAEP's Data Processing Department staff is responsible for providing fast, easy access to the data on a daily basis. They also engage in long-range planning and developmental work and provide general technical support to all the project departments and their advisory committees.

Computer Hardware and Languages

National Assessment uses an IBM 360 Model 20 computer located in Denver, an IBM 360 Model 91 computer located at the Princeton University Computer Center and a high-speed communications link between the two computers. Card decks are read into an IBM 2560 card reader in Denver and transmitted to the Model 91 at Princeton for execution. Printouts are transmitted from Princeton to an IBM 1403 Printer in Denver.

Because of the large amount of data collected by NAEP, the National Military Command Information Processing System (NIPS—a Data Base/Data Management System) is used to process NAEP data. It provides a convenient method for solving the data-handling problems associated with high-volume and large-file processing requirements by using executive programs which simplify common operations such as data definition, data storage, maintenance and retrieval; the amount of human effort required to code, test and debug is reduced.

In addition to NIPS, various general-purpose programming languages, chiefly PL/I and FORTRAN are used. These languages are well suited for statistical computations as well as data manipulations on the various files.

Preparation for an Assessment Year

As the data were being collected in the field, files that would hold the data and the procedures that would be used to manipulate them were being designed at National Assessment. MRC and NAEP discussed specifications for the format in which the data would be prepared by MRC and the edit checks that MRC would perform to insure correct, accurate data. To insure compatibility, formats were also determined for data from special scoring.

Approximately three months before the first MRC-produced data tape was scheduled to be completed, file design and coding began for the respondent file. The magnetic tapes containing response data for each age level were received from MRC at six-to-eight-week intervals. The data was loaded onto the NAEP respondent files as it was received. Approximately four weeks after the MRC tape for an age level was received, a tape containing a weighting factor for each respondent was received from RTI, and the weights were merged with the respondent file.[1]

[1] Refer to Chapter 7, Computation of P-Values, for a discussion of weighting.

Organizing the Data

A variety of files used to check the accuracy of the data and implement various steps in the analyses of the data were developed.

Exercise Documentation File

The exercise documentation file contained exercise-level information for each exercise, keyed by learning area, exercise number, overlap and part. The file included information about the parts of each exercise, such as a short description of the exercise part, the correct foil, the response range, specifications for combinations of parts and analysis specifications. The file also contained specifications needed by the retrieval program used to create auxiliary files.

Respondent File

The respondent file was the main source of data used for all analyses. Four respondent files were created from the MRC and RTI tapes—one for each age group. Each file was organized by package-exercise number and contained records keyed by age, package, PSU (primary sampling unit) and individual identification number—each respondent represented by one record. After the respondent data was loaded onto the files, the weighting factors were loaded along with data obtained from the principal's questionnaires or background questionnaires.

Auxiliary Data Files

Since each analysis used only a fraction of the data from the respondent file, smaller auxiliary files were created in order to increase processing efficiency. Each auxiliary file represented one age level for one learning area. Part of the process of creating these files was to produce data fields which contained combinations of original items of data for each respondent. These additional, derived data fields were stored with the original data on the auxiliary files.

After the auxiliary files had been created, preliminary analysis programs computed percentages of individuals choosing each response for each exercise. They were checked for the validity of response codings and the potential usefulness of the combinations of certain responses. If an invalid coding of a response was discovered, MRC was requested to locate the original package in order to correct the error. Corrections of the data on the respondent file were made and revised specifications for analyses were prepared. New auxiliary files were created and final computer programs were written to calculate percentages, standard errors and other statistics.

Summary Files

The statistics calculated from the auxiliary files were stored on exercise statistic summary files. One summary file was created for each learning area and was keyed by age, package, exercise, part and response. Each record contained the statistics which had been computed for a single response. Additional coded information indicating exercises with similar characteristics was also added to this file, to be used later in the computation of summary statistics across groups of exercises.

Various summary analyses were performed on data from those exercises with similar characteristics. For example, the music performance exercises were summarized, and results for the social studies exercises dealing with political knowledge and attitudes were summarized. A variety of data display programs were used to present tables of medians, means, maxima and minima as well as graphic displays of the distributions of certain exercise statistics. Such summaries provided the basis for generalizations about national and subpopulation performances on various types of exercises.

Appendix File

For each learning area an appendix file was created to produce camera-ready printouts of response statistics for NAEP statistical reports. Descriptive text for each exercise

response was coded, keypunched and edited. This was a massive task. For example, music alone had over 17,000 different response possibilities. Exercise-level statistics were extracted from the summary files and combined with the edited text to create the appendix files.

Availability of Data

In addition to statistical and prose reports produced from the various data analyses, National Assessment also produces computer data tapes for the learning areas assessed. These tapes are created from the respondent files and contain the original response data for the exercises which have been released into the public domain in their entirety. In order to protect anonymity of respondents and political units, the tapes were designed so that it is not possible to relate individual responses to particular respondents, schools, school districts or states. However, there is complete documentation of the classification systems and the contents of the tapes so that the data may be used and reanalyzed to suit the needs of researchers and other interested individuals.

CHAPTER 7

DATA ANALYSIS

National Assessment (NAEP) estimates the percentages of individuals at each of four age levels in the nation as a whole and in certain groups who possess various knowledges, understandings, skills and attitudes.

To do so, the Assessment has developed two baseline measures: (1) the p-value, an estimate of the percentage of persons in each of the populations who gave a certain response to a specific exercise, and (2) the Δp-value (delta-p), an estimate of the difference between the performance of a subpopulation and the performance of the total age population. Measures of change will reflect comparisons of these statistics over time.[1]

The most frequently reported p-values are those that estimate the percentage of a total age population or a certain subpopulation that gave an acceptable response to an exercise. However, NAEP computed many p-values in addition to the percentage of acceptable responses. For multiple-choice exercises, percentages were computed for each unacceptable choice, the "I don't know" response and the total percentage of the unacceptable responses as well. Open-ended exercises generally had several categories of acceptable and unacceptable responses. Percentages were computed for each category as well as for the total number of acceptable and unacceptable responses. Not all exercises had clearly definable acceptable and unacceptable criteria; for example, some were survey questions, and estimates were made of the percentages of people who played certain kinds of musical instruments or read certain kinds of science books or magazines.

Many exercises had several parts or required individuals to give more than one response, each of which could be scored acceptable/unacceptable. Additional analyses summarized results over the parts of an exercise. On these multi-part exercises the percentage of individuals giving exactly zero, one, two, three, etc. acceptable responses were reported as well as the percentage of individuals giving at least one, at least two, at least three, etc. acceptable responses.

For some multi-part exercises, analysis of combinations of responses from two or more parts provided additional interesting information. For example, the following combinations were derived: (1) percentages of individuals who agreed or disagreed with a statement and successfully defended their positions; (2) percentages of people who could not read music, but played a musical instrument acceptably; and (3) percentages of respondents who identified social problems, but could not propose solutions for those problems.

Computation of P-Values

The percentages reported by NAEP are weighted estimates of population percentages. The procedures for developing the estimators and their standard errors are related to the sample design. If every individual had, by

[1] A detailed discussion on National Assessment's statistical computation procedures is available in Robert Larson, Wayne Martin, Todd Rogers, Don Searls, Susan Sherman and David Wright, *A Look at the Analysis of National Assessment Data* presented by J. Stanley Ahmann in *Frontiers of Educational Measurement and Information Systems — 1973*, ed. William E. Coffman (Boston: Houghton Mifflin Company, 1973), pp. 89—111.

design, an equal probability of being selected to respond to a NAEP exercise, a simple ratio of the number of successes to the number responding would be a good estimate of the actual percentage of acceptable responses. In that case, each individual in the sample would represent the same number of individuals in the population. However, in Years 03 and 04 the Assessment determined that all 50 states would be represented in the sample. Also, to insure an adequate number of respondents in the sample at the lower end of the socioeconomic scale, respondents were selected at a higher rate from this subpopulation. The "all states" requirement and oversampling of low socioeconomic individuals resulted in individuals in the population having unequal probabilities of being selected. To adjust for this unequal rate for the sample, appropriate weighting of the data was used to produce unbiased estimates of the percentages. The national percentages—p-values for a total age group—were computed using the weighted responses of all respondents in the sample. For example, the estimated percentage of 9-year-olds who answered acceptably is defined as the sum of the weighted number of acceptable responses divided by the sum of the weighted number of responses for all 9-year-olds. Percentages for each reporting group were computed in the same manner using only that group's weighted responses.

ΔP-Values

A Δp-value is the difference between the estimated percentage of a group (for example, the Western region) that provided a particular response to an exercise and the corresponding national percentage:

Western % — National % = Western Δp-value.

If, on a given exercise, a group's percentage is lower than the national percentage, the difference between the two percentages is expressed as a negative number; if a group's percentage of success is higher than the national percentage, the group's difference in performance is a positive number. For example, if on a given exercise 74% of the 13-year-olds in the West gave an acceptable response, but only 68% of the national sample did so, then the West Δp for 13-year-olds would be +6 percentage points.

Reliable Differences

The p-values are estimates of the population percentages, based on the percentages obtained from the sample. Somewhat different percentages would be obtained if a different sample had been chosen. Variation would occur among the percentages for all possible samples that could have been selected for Years 03 and 04. Since only one sample is drawn per year, an estimate of the variability among all possible values for samples of that size must be computed to determine how precise the obtained estimated percentages were. Since results are reported for each exercise, it is desirable that the variance estimation procedures be inexpensive and yet yield unbiased, stable estimates. The jackknife procedure, a computationally simple and relatively inexpensive procedure, was used in Years 03 and 04.[2] The measure of variability computed is the standard error.

When comparing a group with the nation, it becomes important to decide how large the sample estimate of the difference must be in order to be certain that the difference is not zero. The project has adopted a statistical convention that if an obtained difference is at least twice as large as the estimate of its standard error, then the Δp is said to be reliably different from zero. Statistically, this sort of statement has a 95% probability of being correct. Such differences are called statistically significant and are indicated on exhibits in reports by asterisks (*).

Of course, the practical importance of differences to the educational community is not

[2] F. Mosteller and J.W. Tukey, "Data Analysis, Including Statistics," in *Handbook of Social Psychology* (2nd ed.), eds. G. Lindzey and E. Aronson (Reading, Mass.: Addison-Wesley, 1968), pp. 80—203.

BEST COPY AVAILABLE

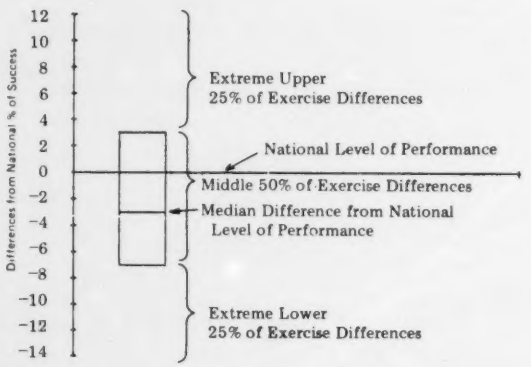
necessarily determined by statistical reliability. The practical importance of larger or smaller differences must be judged by the reader of NAEP reports.

Summary Results

Each learning area is capable of generating a very large number of percentage values. To help readers of NAEP reports, results must be summarized in a general way as well as presented in detail individually. National Assessment summarizes data for clusters of exercises that would be of public interest. For example, in addition to exercises classified as political knowledge and attitudes and those classified as music performance in Year 03, summary analyses will be performed on clusters of math and science exercises.

For a given set of exercises, a group's achievement is summarized conveniently by using its differences from national percentages of success. For example: If on a set of 5 exercises the percentages of success for all 9-year-olds were 90%, 40%, 82%, 75% and 60%, and the percentages of success for a particular group—Northeast 9-year-olds, for instance—on these same exercises were, respectively, 95%, 44%, 85%, 77% and 61%, then the group's differences would be +5%, +4%, +3%, +2% and +1%. The group's median difference from a national performance level—that figure above and below which 50% of its exercise differences lie—would be +3%. If more than 12 exercises are summarized the median difference provides a stable indicator of a group's typical performance over the set of exercises. If one desires a single figure to describe a group's performance relative to a national level of performance on a set of exercises, this is clearly the most useful figure to consult. However, a more complete picture of a group's typical performance emerges from examination of the entire range of differences or, more conveniently, the range of the middle 50% of the exercise differences. For example, given a full range of exercise differences (from +12% to −14% in a hypothetical example) the group's typical performance can be displayed by presenting its median difference and the range of the middle 50% of its exercise differences (see Exhibit 7-1).

EXHIBIT 7-1. How Differences from the National Percentage of Success Are Reported: Sample Graph



If a large proportion of the group's differences occur above (or below) the national level, there is an even stronger indication of how that group's performance differs from the national level. To examine the group's extreme differences, the top 25% and bottom 25% of the distribution of Δp-values can be useful. If a group's performance on any individual exercise deviates considerably from the patterns established in the overall summary data, the deviation could be of special interest.

Results for certain clusters of exercises in each learning area have been compared in various ways. For example, the median difference for exercises clustered together in one reporting topic may have been 8 percentage points below the national level for 13-year-olds in the Southeast. For a different cluster of exercises that group's median difference may have been 5 percentage points above the nation. Therefore, relative to the nation the group tended to perform better on one cluster than on the other.

Estimation of Change

The NAEP procedure for determining change is to reassess certain exercises (approximately half of those assessed in each learning area) four or five years after the previous assessment and compute differences in performance on those exercises. There are at least two straightforward measures of change that will be computed for each exercise:

1. the differences in percentage of correct responses on a given exercise from one assessment to the next and

2. the differences in group Δp-values on a given exercise from one assessment to the next.

Data from the first reassessment (science) are currently being analyzed and will be available in the science change report to be published in 1975.

Limitations of the Data

Within the limitations due to measurement and sampling error, the data computed and reported for Years 03 and 04 accurately describe the educational achievements of the groups designated in the sample.

When the data show that a group's overall level of achievement is either above or below the national level, one must exercise great caution in speculating about the causes. Consider, for example, a hypothetical group whose achievement is well above the national performance level. Most individuals classified as members of the group may attend schools that have excellent physical facilities and high-quality faculties, belong to families that have attained a high socioeconomic level, have well-educated parents and come from homes with many reading materials. Any of these factors could contribute to the group's high level of achievement, while membership in the group itself may contribute very little or nothing. When one looks at the data for a given group, therefore, it cannot be said that any difference in achievement between that group and the nation as a whole is attributable solely to membership in that group. Interpretation of results must take into account the other factors which might be influencing the achievement of the group.

CHAPTER 8

REPORTING AND DISSEMINATING THE FINDINGS

National Assessment (NAEP) generates a great deal of educational information which then must be widely communicated. Given the broad scope of the project and the variety of audiences NAEP serves, this is at best a difficult task.

Various NAEP materials represent different organizations of content and levels of detail. Some publications are highly technical and oriented to the research community; some are less detailed and focus on selective issues or concerns; others, like this yearbook, provide the general background necessary to understand and properly use NAEP results.

At the most technical level are the statistical reports which provide all the information NAEP has gathered in a particular learning area. These reports present each released exercise exactly as it was presented to the respondents and documented to the extent that any secondary user of National Assessment exercises (e.g., a state assessment) can administer it exactly as it was originally administered. For released exercises, percentages appear for the acceptable responses (or categories), unacceptable responses (or categories), the "I don't know" response and no response. Unreleased exercises are neither shown nor completely documented; the data are usually restricted to the correct response, "I don't know" response, no response and all incorrect responses grouped together. All percentages are accompanied by their standard errors. This statistical report, which also contains summary data, is intended primarily for people involved in state and local assessments, educational researchers, educational decisionmakers and any other groups or persons who may need or desire complete and documented data. Other publications of this nature include special monographs on sampling and background variables.

Selective reports, on the other hand, deal with timely, important or interesting aspects of a learning area in a nontechnical way. An overview report in each learning area provides a broad view of the findings in popular language with some specific, interesting details highlighted. These reports are intended primarily for professional associations, subject-matter groups, civic groups, public officials at local, state and national levels and those educational decisionmakers and any other persons or groups who may neither need nor desire complete and documented data.

The Assessment does not attempt to interpret its findings or discuss their implications for the educational community in any of its reports. Under the auspices of National Assessment, however, various professional groups or panels of professional individuals are writing interpretive reports. For instance, the National Science Teachers Association has published an interpretive study of the results of the first science assessment, and the National Council of Teachers of Mathematics plans to do the same in 1974–75. The interpretations, implications and evaluations in these reports are solely those of the professional groups or panels providing them and do not reflect NAEP's point of view.

Finally, the project's findings appear in its *Newsletter*, in press releases and newspaper columns, in the publications of professional groups such as National Council of Teachers of English, Music Educators National Conference or National Council of Teachers of Mathematics, in general education publications such as *American Education* or *Compact* and even in magazines such as *Reader's Digest*.

CHAPTER 9

ADAPTATION AND UTILIZATION OF THE NATIONAL ASSESSMENT MODEL

Although data is gathered on a national level, National Assessment (NAEP) recognizes the potential importance of its model, technology and materials as they relate to the solution of problems at state and local levels. For this reason, NAEP has created the Model Utilization/Adaptation (MU/A) program to encourage and facilitate the adaptation of the Assessment model, technology and materials (or selected aspects of these) by state and local educational systems. To the extent that this is done, questions regarding the educational outcomes being answered at the national level can be answered in a more refined sense at state and local levels. The MU/A program provides services to state and local educational systems in three ways:

1. consultation between MU/A staff and state or local officials to determine the feasibility of adapting the NAEP model to meet local needs,

2. technical assistance to state and local officials in the use of the Assessment model or certain aspects of it and

3. workshops and seminars that provide an exchange of ideas between NAEP and state and local assessments.

Because the project does not presently have the resources to deal with the many diverse requests from local educational systems, these MU/A services can be made available to officials at the local level only on a limited basis unless special circumstances dictate otherwise. Generally, the MU/A program does no more than provide available materials to local educational systems and communicate with them by telephone or mail about general problems, referring them to their state departments of education for information about content-referenced assessment. In certain instances, further consultation, technical assistance and participation in workshops and seminars are made available to local educational systems:

1. if the local assessment is an integral part of a state assessment plan involving the adaptation of NAEP procedures and

2. if NAEP foresees the possibility for the demonstration of a new form of adaptability that could provide valuable insights into significant uses of NAEP procedures at the local level.

The MU/A program provides consultation to officials not directly concerned with assessment through a state educational system or a local educational system only when a clear and distinct possibility of improving educational decision making can be established in preliminary discussions.

Consultation

Upon request from state officials, the MU/A program provides consultation to state educational agencies, state boards of education, state legislative groups and statewide citizen groups.

In providing consultation through its MU/A program, National Assessment does not actually implement state (or local) assessments, provide prefabricated assessment plans nor endorse any state (or local) assessment. Consultation is provided as input to assist an

educational agency in determining the feasibility of adapting the NAEP model, technology or materials to the agency's specific assessment needs.

Some of the questions that might be discussed by the MU/A staff and state officials are:

1. What reporting categories best fit the state's needs?

2. If sampling is to be used, what sample design is most efficient both in terms of cost and obtaining the requisite data for the reporting categories?

3. What analyses best fit the sample design and reporting needs?

4. Should the state use all of NAEP's exercises, just the easily administered and easily scored multiple-choice exercises, just those pertaining to certain objectives or just a portion of those pertaining to each objective?

Technical Assistance

If the MU/A consultation with state officials has been fruitful and the state has decided to adapt the NAEP model, technologies or materials for its assessment, the MU/A program provides continued technical assistance. Once a state has decided that it is feasible to adapt some aspect(s) of the project, the MU/A program, through technical assistance, helps the state to inaugurate and carry out the adaptations to assure the best possible chance of success.

Technical assistance is generally provided directly by the MU/A staff. However, depending upon the nature and degree of expertise required, other NAEP staff or outside consultants may be required to provide the assistance. For example, a state particularly concerned about the implications of sampling may require the technical assistance from the NAEP sampling coordinator.

Workshops and Seminars

Because the MU/A program is relatively new and because the entire area of content-referenced assessment at state and local levels is also new and constantly evolving, the program relies heavily upon the dissemination and exchange of ideas through personal contact. To facilitate this dissemination and exchange of ideas, NAEP conducts an annual assessment workshop. Most of the participants represent state educational agencies.

Currently, two types of workshops are being conducted. One type is for participants who are relatively new to the field of educational assessment. Its primary purpose is to introduce approaches developed by NAEP. Examples of some of the topics covered in these workshops include discussions on how the NAEP model can be adapted for the types of data that might be provided for educational decision making. This allows the participants to examine the ways in which such a model could be adapted to the needs of their various states. This workshop also provides the participants with an opportunity to become aware of other services (consultation and technical assistance) the project makes available to states.

The second type is a conference workshop for those already exposed to educational assessment. This conference allows the participants to discuss their experiences with state assessment during the year, to gain insight from the experiences of other states and to learn of any changes in the NAEP model.

Specialized seminars concerning sampling and analysis, reporting disseminating and utilizing and applying assessment data are anticipated. Previous experiences have shown these topics to be of great interest to state department personnel planning and conducting assessments. These seminars will be limited to participants involved in the sampling/analysis or reporting/dissemination/application phases of assessment in order that more detailed information and personal interaction can be provided.

As state and local educational agencies develop and implement assessment plans adapting the Assessment model, technologies and procedures, the MU/A program develops materials to fulfill two purposes:

1. documentation of the plans and methodologies for implementation, including both success and problem areas, for those who plan later assessments and
2. documentation of the utility of analyzed and interpreted data from state and local assessments to show other educational agencies the usefulness of assessments based upon adaptations of various components of the NAEP model.

BEST COPY AVAILABLE

APPENDIX A

DEFINITIONS OF NATIONAL ASSESSMENT REPORTING POPULATIONS

National Assessment (NAEP) reporting populations include 9, 13 and 17-year-olds enrolled in public or private schools as well as 17-year-olds not enrolled in school and young adults aged 26 to 35. Within these age groups, results are reported by geographic region, sex, color, parental education and size and type of community.

With a few exceptions, listed below, the reporting populations include all age-eligible persons in the 50 states and the District of Columbia. Nine and 13-year-old populations are restricted to those students enrolled in public or private schools during the survey period. The 17-year-old population includes students enrolled in public or private schools during the survey period, those age-eligible persons not enrolled but identified by schools as having dropped out or graduated early or those age-eligible persons not enrolled but living in identifiable housing units in the United States. Survey periods and age ranges for the four age groups are shown in Exhibit A-1. Age-eligible persons are excluded if they are:

1. non-English speaking,
2. institutionalized and
3. physically, emotionally, or mentally handicapped in such a way that they cannot respond to the exercises as administered.

EXHIBIT A-1. National Assessment Populations, Survey Periods and Age Ranges

Age Level	Year 03 Survey Period*	Year 03 Birth-Date Range*
9	January – February 1972	1/1/62 – 12/31/62
13	October – December 1971	1/1/58 – 12/31/58
17 in-school	March – April 1972	10/1/54 – 9/30/55
17 not enrolled in school during January 1972	January – July 1972	10/1/54 – 9/30/55
17 not enrolled in school during March 1971	January – July 1972	10/1/53 – 9/30/54
Young adults (Year 03)	January – July 1972	4/1/36 – 3/31/46
Young adults (Year 04)	October 1972 – May 1973	1/1/37 – 12/31/46

**For Year 04 survey periods and birth-date ranges of 9, 13 and 17-year-olds, add one year to the Year 03 dates.*

National Assessment Groups

Geographic Regions

Results are reported by the four geographic regions defined by the Office of Business Economics, Department of Commerce. The states in each region are shown in Exhibit A-2.

EXHIBIT A-2. Definitions of National Assessment Regional Subpopulations*

Northeast	Southeast
Delaware	Arkansas
Connecticut	Florida
Maine	Virginia
New Hampshire	West Virginia
Rhode Island	Alabama
Vermont	Georgia
District of Columbia	Kentucky
Maryland	Louisiana
Massachusetts	Mississippi
New Jersey	North Carolina
Pennsylvania	South Carolina
New York	Tennessee
Central	**West**
Iowa	Alaska
Kansas	Hawaii
Nebraska	Idaho
North Dakota	Montana
South Dakota	Nevada
Minnesota	Wyoming
Missouri	Arizona
Illinois	Oregon
Indiana	Utah
Michigan	Colorado
Wisconsin	New Mexico
Ohio	Oklahoma
	California
	Texas
	Washington

**These regional subpopulation definitions are the same as those used by the Office of Business Economics, Department of Commerce.*

Sex

Results are reported separately for males and females at all age levels.

Color

Results are reported for Whites and Blacks. While respondents were classified into several other ethnic groups, including Puerto Ricans, Mexican-American and other or unclassified, the actual sample sizes realized were insufficient for reporting purposes.

Parental Education

Parental education refers to the highest level of education level reported by the respondent for either parent.

No high school. Neither parent has any formal education beyond the eighth grade.

Some high school. At least one parent has some formal education beyond the eighth grade, but neither parent has graduated from high school.

Graduated from high school. At least one parent has graduated from high school, but neither parent has any formal education beyond high school.

Post high school. At least one parent has some formal education beyond high school including any business, professional or trade school training as well as college or university training.

Size and Type of Community (STOC)

The seven size and type of community (STOC) reporting categories are comprised of three "extreme" types of community (TOC) and four "residual" sizes of community (SOC). Each TOC category includes approximately 10% of the respondents at each age level; the remaining respondents are classified according to one of the SOC classifications.

Briefly, the three TOC categories are: (1) city areas where a high proportion of the adult population is either not regularly employed or

on welfare and a low proportion is employed in professional or managerial positions; (2) rural areas where a high proportion of adults are farm workers and a low proportion are professional, managerial or factory workers; and (3) near-city and city areas where a high proportion of adults are employed in professional or managerial positions and a low proportion are factory or farm workers, not regularly employed or on welfare. Respondents are placed in one of these categories if the occupational profile and location of the school or, in the case of the out-of-school sample, segment satisfy the extreme TOC definitions.

The remaining respondents at each age level are classified according to the size of community in which the school or segment is located. The occupational profile is based on the employment categories summarized in Exhibit A-3.

For the in-school sample at each age and the supplementary sample at age 17, the school principal of each selected school provided estimates of the percentage of students whose parents fit into each occupational category. Other occupational data for the out-of-school sample was obtained from census data in Year 03 and from the respondents themselves in Year 04. The definitions used to classify respondents by STOC are presented in Exhibit A-4. The occupational index is computed using the occupational categories summarized in Exhibit A-3. The distribution of respondents by STOC category for Year 03 is presented in Exhibit A-5.

EXHIBIT A-3. Occupational Categories

Categories	Code
Professional or managerial personnel	A
Sales, clerical, technical or skilled workers	B
Factory or other blue collar workers	C
Farm workers	D
Not regularly employed	E
On welfare	F

EXHIBIT A-4. National Assessment Size and Type of Community (STOC) Reporting Categories

Reporting Category	Occupational Index*	Description
Low metro	E+F-A	Sample schools or segments in a city or metropolitan area of a city with a population greater than 150,000 and in the 90–99th percentile of the low metro index
Extreme rural	D-(C+2A)	Sample schools or segments in communities with a population less than 8,000 and in the 90–99th percentile of the extreme rural index
Small place		Sample schools or segments in a community with a population less than 25,000 and not classified as extreme rural
Medium city		Sample schools or segments in a city with a population between 25,000 and 200,000 and not classified as low metro or high metro
Main big city		Sample schools or segments within the city limits of a city with a population greater than 200,000 and not class fied as high metro or low metro
Urban fringe		Sample schools or segments in the metropolitan area of a big city but outside the city limits and not classified as low metro, extreme rural or high metro
High metro	A-(C+D+E+F)	Sample schools or segments in a city or metropolitan area of a city with a population greater than 150,000 and in the 90–99th percentile on the high metro index

**See Exhibit A-3.*

EXHIBIT A-5. Distribution of Respondents by Size and Type of Community (STOC) and Age Level

STOC Category	Age Level					
	9	13	17 (IS)*	17 (OS)†	17 (All)	Adult
1. Low metro	10.2%	9.9%	11.3%	14.0%	11.7%	9.9%
2. Extreme rural	10.0	9.6	11.8	9.8	11.6	10.2
3. Small place	32.4	33.7	33.0	33.4	33.1	31.8
4. Medium city	16.8	16.6	15.9	19.9	16.3	13.7
5. Main big city	10.5	11.2	9.3	6.4	9.0	8.0
6. Urban fringe	10.0	8.4	5.1	7.7	5.4	16.4
7. High metro	10.1	10.7	13.6	8.9	13.0	10.0

**In-school*
†Out-of-school

APPENDIX B

ESTIMATED POPULATION DISTRIBUTIONS OF NATIONAL ASSESSMENT REPORTING GROUPS

Based on Year 03 data, Exhibits B-1 through B-4 contain the percentage of the population at each age level estimated to be in each two-way combination of the reporting groups defined in Appendix A. These data provide a more complete description of the composition of the population at each age level in terms of National Assessment's (NAEP) reporting variables. Each entry is an estimate of the percent of a reporting group identified by a row heading found in a second reporting group identified by a column heading. For example, in each exhibit, the rows corresponding to color—Black, White, Other—show the estimated percentages of each color group in the nation, each region, sex, parental education level and age and type of community classification. As shown in Exhibit B-1, the estimated percentage of 9-year-old Blacks (row 7) in the nation (column 1) is 13.6%, while of the total population of 9-year-olds in the Southeast (column 2), 26.6% are Black (row 7). The estimated percentage of 9-year-old Blacks living in the Southeast, 45.5%, is shown in row 1, column 8. The exhibits have been arranged so that for each reporting group identified by a column heading the percentages across the levels of a reporting variable identified by rows add to 100%. For example, the sum of the percentage of Blacks, Whites and Other equal 100% in every column of the exhibits.

BEST COPY AVAILABLE

EXHIBIT B-1a

Estimated Population Distributions of National Assessment Reporting Groups (9-Year-Olds)

Variables and Groups	Natl. (%)	Region (%)				Sex (%)		Color (%)		
		SE	W	C	NE	M	F	B	W	Oth.
Region										
Southeast	23.2	100.0	0.0	0.0	0.0	22.7	23.7	45.5	21.1	1.2
West	24.0	0.0	100.0	0.0	0.0	23.5	24.5	18.9	20.5	81.2
Central	27.8	0.0	0.0	100.0	0.0	28.2	27.4	18.8	31.2	3.9
Northeast	25.1	0.0	0.0	0.0	100.0	25.7	24.5	16.8	27.4	13.7
Sex										
Male	49.1	47.9	47.9	49.7	50.2	100.0	0.0	47.0	49.5	47.5
Female	51.1	52.1	52.1	50.3	49.8	0.0	100.0	53.0	50.6	52.5
Color										
Black	13.6	26.6	10.7	9.2	9.1	13.0	14.1	100.0	0.0	0.0
White	80.4	73.1	68.6	90.0	87.6	81.1	79.7	0.0	100.0	0.0
Other	6.2	0.3	20.8	0.9	3.4	6.0	6.3	0.0	0.0	100.0
Parental education										
No high school	5.5	9.5	6.0	4.2	2.8	5.4	5.5	7.2	4.6	13.8
Some high school	4.9	7.5	4.1	5.0	3.0	5.0	4.7	6.7	4.5	5.6
Graduated high school	23.4	25.3	23.0	25.4	19.7	24.6	22.3	23.7	23.7	18.8
Post high school	35.2	30.3	38.5	33.9	37.7	36.7	33.6	27.1	37.6	20.7
Unknown	31.3	27.5	28.4	31.6	36.8	28.3	33.9	35.3	29.7	41.0
Size and type of community (STOC)										
Low metro	8.9	9.4	11.9	8.5	5.7	8.5	9.2	44.5	1.8	22.3
Extreme rural	8.6	10.7	10.5	11.2	1.7	8.7	8.4	5.6	8.8	12.3
Small place	29.7	36.4	23.1	27.0	32.6	29.3	30.1	15.3	32.6	23.3
Medium city	16.8	24.2	15.6	19.8	7.8	17.0	16.6	14.5	17.6	11.6
Main big city	10.5	5.9	9.9	11.0	14.5	10.2	10.6	12.6	9.7	15.9
Urban fringe	11.5	6.7	13.9	10.7	14.4	11.8	11.2	4.9	12.7	9.3
High metro	14.3	6.8	15.1	11.7	23.3	14.7	13.9	2.7	17.0	5.3

BEST COPY AVAILABLE

EXHIBIT B-1b

Estimated Population Distributions of National Assessment Reporting Groups (9-Year-Olds)

Variables and Groups	Parental Education (%)					Size and Type of Community (%)						
	NHS	SHS	GHS	PHS	Unk.	LM	ER	SP	MC	MBC	UF	HM
Region												
Southeast	40.0	35.9	25.1	20.0	20.4	24.6	29.0	28.4	33.4	13.1	13.5	11.1
West	26.2	20.1	23.6	26.3	21.9	32.4	29.4	18.7	22.3	22.8	29.1	25.4
Central	21.1	28.4	30.2	26.8	28.2	26.7	36.5	25.3	32.8	29.4	26.0	22.8
Northeast	12.8	15.7	21.1	26.9	29.6	16.3	5.1	27.6	11.6	34.8	31.5	40.8
Sex												
Male	48.6	50.5	51.5	51.2	44.5	47.0	49.7	48.3	49.6	48.0	50.2	50.3
Female	51.4	49.5	48.5	48.8	55.5	53.0	50.3	51.7	50.5	52.0	49.8	49.7
Color												
Black	17.8	18.8	13.7	10.5	15.3	68.2	8.8	7.0	11.6	16.3	5.8	2.5
White	66.8	74.1	81.4	85.9	76.7	16.3	82.3	88.2	84.2	74.4	89.2	95.2
Other	15.5	7.1	4.9	3.6	8.1	15.5	8.9	4.8	4.2	9.4	5.0	2.3
Parental education												
No high school	100.0	0.0	0.0	0.0	0.0	7.2	10.4	6.8	5.6	2.9	4.2	1.5
Some high school	0.0	100.0	0.0	0.0	0.0	5.3	6.7	6.0	5.2	3.3	4.5	1.9
Graduated high school	0.0	0.0	100.0	0.0	0.0	21.3	28.2	26.9	23.7	19.9	23.7	16.4
Post high school	0.0	0.0	0.0	100.0	0.0	27.2	24.8	32.5	34.6	31.4	34.4	55.5
Unknown	0.0	0.0	0.0	0.0	100.0	38.9	29.9	27.8	30.8	42.5	33.2	24.8
Size and type of community (STOC)												
Low metro	11.6	9.7	8.0	6.9	11.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Extreme rural	16.2	11.8	10.3	6.0	8.2	0.0	100.0	0.0	0.0	0.0	0.0	0.0
Small place	36.9	36.9	34.1	27.4	26.4	0.0	0.0	100.0	0.0	0.0	0.0	0.0
Medium city	17.3	18.2	17.0	16.6	16.6	0.0	0.0	0.0	100.0	0.0	0.0	0.0
Main big city	5.4	7.2	8.9	9.3	14.2	0.0	0.0	0.0	0.0	100.0	0.0	0.0
Urban fringe	8.8	10.7	11.6	11.2	12.2	0.0	0.0	0.0	0.0	0.0	100.0	0.0
High metro	3.8	5.6	10.0	22.6	11.3	0.0	0.0	0.0	0.0	0.0	0.0	100.0

EXHIBIT B-2a

Estimated Population Distributions of National Assessment Reporting Groups (13-Year-Olds)

Variables and Groups	Natl. (%)	Region (%)				Sex (%)		Color (%)		
		SE	W	C	NE	M	F	B	W	Oth.
Region										
Southeast	22.6	100.0	0.0	0.0	0.0	22.8	22.4	46.1	21.1	1.9
West	24.1	0.0	100.0	0.0	0.0	24.1	24.1	12.7	21.7	79.3
Central	29.4	0.0	0.0	100.0	0.0	29.3	29.5	23.8	31.8	5.0
Northeast	24.1	0.0	0.0	0.0	100.0	24.0	24.2	17.5	25.6	13.9
Sex										
Male	50.0	50.5	49.9	49.9	49.7	100.0	0.0	46.7	50.5	49.5
Female	50.1	49.5	50.1	50.2	50.3	0.0	100.0	53.3	49.7	50.5
Color										
Black	10.3	21.1	5.4	8.4	7.5	9.7	11.0	100.0	0.0	0.0
White	83.9	78.5	75.6	90.7	89.2	84.7	83.2	0.0	100.0	0.0
Other	5.8	0.5	19.0	1.0	3.3	5.7	5.8	0.0	0.0	100.0
Parental education										
No high school	6.5	10.2	7.4	5.4	3.4	6.4	6.5	9.8	5.1	19.9
Some high school	9.2	14.0	8.3	8.7	6.1	8.2	10.2	16.2	8.2	11.2
Graduated high schoo'	31.5	30.7	25.5	37.6	30.4	31.4	31.5	29.1	32.5	20.9
Post high school	41.1	33.8	46.5	38.8	45.2	41.6	40.6	21.9	44.8	21.1
Unknown	11.9	11.3	12.3	9.6	15.0	12.4	11.2	23.0	9.4	26.9
Size and type of community (STOC)										
Low metro	7.3	6.0	8.5	8.1	6.1	7.2	7.3	26.9	3.6	25.1
Extreme rural	8.3	10.9	6.7	13.1	1.6	8.2	8.5	8.3	8.4	6.8
Small place	32.4	40.4	26.7	31.8	30.8	33.0	31.7	23.8	34.1	22.7
Medium city	16.2	22.2	15.5	16.0	11.2	16.0	16.3	14.2	16.6	13.1
Main big city	11.6	10.5	7.9	16.0	10.7	11.4	11.7	17.6	10.7	13.2
Urban fringe	10.9	2.9	17.3	5.4	18.4	10.9	10.8	4.4	11.5	12.8
High metro	13.8	7.2	17.4	9.6	21.?	13.6	13.9	4.8	15.4	6.4

BEST COPY AVAILABLE

EXHIBIT B-2b

Estimated Population Distributions of National Assessment Reporting Groups (13-Year-Olds)

Variables and Groups	Parental Education (%)					Size and Type of Community (%)						
	NHS	SHS	GHS	PHS	Unk.	LM	ER	SP	MC	MBC	UF	HM
Region												
Southeast	35.5	34.5	22.0	18.6	21.5	18.7	29.5	28.2	31.0	20.5	6.1	11.8
West	27.6	21.9	19.6	27.2	24.7	28.3	19.5	19.9	23.2	16.4	38.3	30.5
Central	24.3	27.8	35.2	27.7	25.6	32.7	46.4	28.9	29.1	40.8	14.7	20.5
Northeast	12.7	15.9	23.3	26.5	30.3	20.3	4.6	23.0	16.8	22.4	40.9	37.2
Sex												
Male	49.7	44.5	49.8	50.7	52.5	49.6	49.0	51.0	49.4	49.2	50.2	49.4
Female	50.3	55.5	50.2	49.4	47.6	50.4	51.0	49.0	50.6	50.8	49.8	50.6
Color												
Black	15.7	18.2	9.6	5.5	19.9	38.4	10.3	7.6	9.1	15.8	4.2	3.6
White	66.6	74.7	86.6	91.5	67.1	41.7	85.0	88.4	86.2	77.7	89.0	93.7
Other	17.7	7.1	3.8	3.0	13.0	20.0	4.7	4.1	4.7	6.6	6.8	2.7
Parental education												
No high school	100.0	0.0	0.0	0.0	0.0	10.5	14.5	7.5	5.5	5.8	3.5	1.0
Some high school	0.0	100.0	0.0	0.0	0.0	11.4	12.8	11.1	9.1	9.8	6.9	2.5
Graduated high school	0.0	0.0	100.0	0.0	0.0	29.3	35.6	34.8	30.9	32.5	33.7	20.2
Post high school	0.0	0.0	0.0	100.0	0.0	23.5	26.1	36.8	43.4	33.2	45.2	69.8
Unknown	0.0	0.0	0.0	0.0	100.0	25.4	11.1	9.8	11.2	18.7	10.7	6.5
Size and type of community (STOC)												
Low metro	11.7	9.0	6.8	4.1	15.4	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Extreme rural	18.6	11.6	9.4	5.3	7.7	0.0	100.0	0.0	0.0	0.0	0.0	0.0
Small place	37.6	39.3	35.7	29.0	26.5	0.0	0.0	100.0	0.0	0.0	0.0	0.0
Medium city	13.8	15.9	15.8	17.1	15.1	0.0	0.0	0.0	100.0	0.0	0.0	0.0
Main big city	10.3	12.3	11.9	9.3	18.1	0.0	0.0	0.0	0.0	100.0	0.0	0.0
Urban fringe	5.9	8.2	11.6	11.9	9.7	0.0	0.0	0.0	0.0	0.0	100.0	0.0
High metro	2.1	3.8	8.8	23.4	7.5	0.0	0.0	0.0	0.0	0.0	0.0	100.0

EXHIBIT B-3a

Estimated Population Distributions of National Assessment Reporting Groups (All 17-Year-Olds)

Variables and Groups	Natl. (%)	Region (%)				Sex (%)		Color (%)		
		SE	W	C	NE	M	F	B	W	Oth.
Region										
Southeast	21.6	100.0	0.0	0.0	0.0	21.1	22.2	42.9	20.1	0.8
West	24.9	0.0	100.0	0.0	0.0	24.9	25.0	18.5	22.2	80.6
Central	28.2	0.0	0.0	100.0	0.0	28.7	27.8	21.9	30.6	4.8
Northeast	25.2	0.0	0.0	0.0	100.0	25.4	25.1	16.6	27.1	13.9
Sex										
Male	48.7	47.4	48.6	49.5	49.1	100.0	0.0	46.1	49.0	49.3
Female	51.3	52.6	51.4	50.5	51.0	0.0	100.0	53.9	51.0	50.7
Color										
Black	11.3	22.3	8.4	8.8	7.4	10.7	11.8	100.0	0.0	0.0
White	83.3	77.5	74.2	90.3	89.6	83.9	82.8	0.0	100.0	0.0
Other	5.4	0.2	17.5	0.9	3.0	5.5	5.3	0.0	0.0	100.0
Parental education										
No high school	8.4	12.3	9.7	6.8	5.7	8.2	8.7	15.1	6.1	29.8
Some high school	12.6	16.9	11.8	12.2	10.1	11.3	13.9	22.7	10.9	17.4
Graduated high school	30.3	28.5	25.7	35.2	30.8	30.9	29.7	27.8	31.4	18.8
Post high school	43.6	38.4	46.5	42.2	46.9	44.0	43.3	23.8	47.8	20.5
Unknown	5.1	3.8	6.3	3.6	6.5	5.6	4.5	10.6	3.8	13.5
Size and type of community (STOC)										
Low metro	10.7	7.1	14.1	10.9	10.2	10.7	10.7	38.0	5.9	28.4
Extreme rural	9.2	11.3	6.1	15.5	3.3	9.2	9.1	6.6	9.4	10.5
Small place	31.0	36.2	29.4	30.7	28.3	30.5	31.4	19.5	33.2	20.7
Medium city	16.5	23.2	15.6	15.3	13.0	16.5	16.5	16.9	16.7	9.2
Main big city	9.3	8.5	11.0	7.2	10.8	8.9	9.8	12.4	8.3	18.4
Urban fringe	7.2	3.5	9.9	7.1	7.8	6.8	7.5	1.6	8.0	6.2
High metro	16.2	10.3	14.1	13.3	26.6	17.4	15.0	5.2	18.3	6.6

BEST COPY AVAILABLE

EXHIBIT B-3b

Estimated Population Distributions of National Assessment Reporting Groups (All 17-Year-Olds)

Variables and Groups	Parental Education (%)					Size and Type of Community (%)						
	NHS	SHS	GHS	PHS	Unk.	LM	ER	SP	MC	MBC	UF	HM
Region												
Southeast	31.5	29.0	20.4	19.1	16.4	14.3	26.7	25.3	30.4	19.6	10.4	13.7
West	28.8	23.3	21.1	26.6	30.9	32.8	16.5	23.7	23.6	29.3	34.3	21.7
Central	22.6	27.3	32.8	27.3	20.3	28.8	47.7	28.0	26.1	21.8	27.9	23.2
Northeast	17.1	20.3	25.7	27.1	32.4	24.1	9.2	23.0	19.9	29.3	27.4	41.4
Sex												
Male	47.4	43.6	49.7	49.1	54.5	48.8	49.0	48.0	48.7	46.3	46.3	52.4
Female	52.6	56.4	50.3	50.9	45.5	51.2	51.0	52.0	51.3	53.7	53.7	47.6
Color												
Black	20.2	20.3	10.3	6.1	23.7	40.0	8.1	7.1	11.4	15.0	2.5	3.6
White	60.7	72.3	86.3	91.3	62.0	45.7	85.7	89.3	85.6	74.3	92.8	94.2
Other	19.1	7.5	3.4	2.5	14.4	14.4	6.2	3.6	3.0	10.6	4.7	2.2
Parental education												
No high school	100.0	0.0	0.0	0.0	0.0	10.3	14.9	9.9	7.4	7.5	8.4	2.4
Some high school	0.0	100.0	0.0	0.0	0.0	16.4	15.6	15.1	12.2	12.5	11.2	4.7
Graduated high schoo	0.0	0.0	100.0	0.0	0.0	30.9	35.4	33.1	31.1	28.8	28.7	22.1
Post high school	0.0	0.0	0.0	100.0	0.0	32.9	29.4	38.6	45.5	36.2	48.3	68.8
Unknown	0.0	0.0	0.0	0.0	100.0	9.5	4.7	3.3	3.8	14.9	3.5	2.0
Size and type of community (STOC)												
Low metro	13.0	13.9	10.9	8.1	20.2	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Extreme rural	16.2	11.3	10.7	6.2	8.6	0.0	100.0	0.0	0.0	0.0	0.0	0.0
Small place	36.3	37.1	33.9	27.4	20.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
Medium city	14.5	15.9	16.9	17.2	12.3	0.0	0.0	0.0	100.0	0.0	0.0	0.0
Main big city	8.3	9.3	8.9	7.7	27.5	0.0	0.0	0.0	0.0	100.0	0.0	0.0
Urban fringe	7.1	6.4	6.8	8.0	5.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
High metro	4.5	6.1	11.8	25.5	6.4	0.0	0.0	0.0	0.0	0.0	0.0	100.0

EXHIBIT B-4a

Estimated Population Distributions of National Assessment Reporting Groups (Young Adults)

Variables and Groups	Natl. (%)	Region (%)				Sex (%)		Color (%)		
		SE	W	C	NE	M	F	B	W	Oth.
Region										
Southeast	21.2	100.0	0.0	0.0	0.0	21.2	21.2	34.7	20.6	4.6
West	25.5	0.0	100.0	0.0	0.0	25.7	25.3	15.8	23.7	71.4
Central	29.0	0.0	0.0	100.0	0.0	29.1	29.0	22.4	31.4	5.6
Northeast	24.2	0.0	0.0	0.0	100.0	23.9	24.5	27.1	24.3	18.3
Sex										
Male	49.0	49.0	49.5	49.1	48.4	100.0	0.0	45.9	49.1	53.6
Female	51.0	51.0	50.5	50.9	51.6	0.0	100.0	54.2	50.9	46.4
Color										
Black	10.7	17.4	6.6	8.3	12.0	10.0	11.4	100.0	0.0	0.0
White	83.7	81.3	77.8	90.7	83.8	83.9	83.6	0.0	100.0	0.0
Other	5.6	1.2	15.6	1.1	4.2	6.1	5.1	0.0	0.0	100.0
Parental education										
No high school	28.4	36.3	26.0	28.1	24.5	27.7	29.1	41.8	24.4	63.4
Some high school	15.9	16.5	13.1	17.3	16.7	15.1	16.7	22.6	15.6	8.6
Graduated high school	29.8	23.2	32.1	30.4	32.6	31.5	28.2	21.4	31.9	15.0
Post high school	22.4	18.9	27.2	20.2	23.2	22.5	22.4	7.2	25.2	9.8
Unknown	3.4	5.1	1.6	4.0	3.0	3.2	3.5	7.0	2.9	3.2
Size and type of community (STOC)										
Low metro	9.0	3.5	11.3	9.0	11.4	9.1	9.0	42.1	3.9	22.8
Extreme rural	7.6	7.7	6.3	14.7	0.5	7.6	7.7	6.2	7.7	9.8
Small place	30.4	43.2	29.7	27.0	24.0	30.8	30.0	13.5	33.6	14.8
Medium city	12.8	26.5	10.3	10.3	6.4	13.1	12.5	16.4	12.8	6.4
Main big city	10.5	2.0	16.8	8.6	13.6	10.6	10.4	12.0	8.9	31.2
Urban fringe	17.5	6.9	13.8	20.9	26.7	17.7	17.3	6.8	19.6	7.4
High metro	12.2	10.2	11.9	9.5	17.4	11.2	13.2	3.1	13.6	7.8

BEST COPY AVAILABLE

EXHIBIT B-4b

Estimated Population Distributions of National Assessment Reporting Groups (Young Adults)

Variables and Groups	Parental Education (%)					Size and Type of Community (%)						
	NHS	SHS	GHS	PHS	Unk.	LM	ER	SP	MC	MBC	UF	HM
Region												
Southeast	27.1	22.0	16.5	17.9	32.1	8.4	21.5	30.2	44.0	4.1	8.3	17.7
West	23.3	21.0	27.5	30.9	12.1	32.0	20.9	24.9	20.5	40.8	20.1	24.9
Central	28.7	31.5	29.6	26.2	34.3	29.1	55.9	25.8	23.4	23.8	34.7	22.7
Northeast	20.9	25.5	26.4	25.1	21.5	30.6	1.7	19.1	12.1	31.3	36.9	34.7
Sex												
Male	47.8	46.4	51.8	49.1	46.9	49.3	48.6	49.7	50.3	49.5	49.6	44.9
Female	52.2	53.6	48.2	50.9	53.1	50.7	51.4	50.3	49.7	50.5	50.4	55.1
Color												
Black	15.7	15 2	7.7	3.5	22.1	49.9	8.7	4.7	13.7	12.2	4.1	2.7
White	71.9	81.8	89.5	94.1	72.7	36.0	84.2	92.6	83.6	71.2	93.5	93.7
Other	12.4	3.0	2.8	2.4	5.3	14.1	7.1	2.7	2.8	16.6	2.4	3.6
Parental education												
No high school	100.0	0.0	0.0	0.0	0.0	37.0	40.9	28.9	31.0	33.3	25.0	11.4
Some high school	0.0	100.0	0.0	0.0	0.0	24.7	13.8	14.7	17.3	17.4	16.4	10.3
Graduated high school	0.0	0.0	100.0	0.0	0.0	24.2	25.5	29.8	27.9	26.5	35.4	33.7
Post high school	0.0	0.0	0.0	100.0	0.0	8.5	16.2	21.8	20.7	21.5	20.1	44.2
Unknown	0.0	0.0	0.0	0.0	100.0	5.5	3.6	4.8	3.2	1.3	3.2	0.4
Size and type of community (STOC)												
Low metro	11.7	14.0	7.3	3.4	14.7	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Extreme rural	11.0	6.6	6.5	5.5	8.2	0.0	100.0	0.0	0.0	0.0	0.0	0.0
Small place	30.9	28.1	30.4	29.6	42.8	0.0	0.0	100.0	0.0	0.0	0.0	0.0
Medium city	13.9	13.9	12.0	11.8	12.2	0.0	0.0	0.0	100.0	0.0	0.0	0.0
Main big city	12.3	11.5	9.4	10.1	4.1	0.0	0.0	0.0	0.0	100.0	0.0	0.0
Urban fringe	15.4	18.1	20.7	15.7	16.6	0.0	0.0	0.0	0.0	0.0	100.0	0.0
High metro	4.9	7.9	13.8	24.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0	100.0

APPENDIX C

NATIONAL ASSESSMENT POLICY COMMITTEE

(Chairman of Policy Committee)
Mr. Robert McBride, President-Elect
Nat'l. Assoc. of School Boards of Education

Dr. George Brain
Dean, College of Education
Washington State University
Pullman, Washington

Mr. William R. Conway
Education Committee
Florida House of Representatives

Dr. Lyman V. Ginger
Superintendent of Public Instruction
State Department of Education
Frankfort, Kentucky

Dr. George Kozmetsky
Dean, Graduate School of Business
University of Texas

Ms. Joyce E. Lewis
Maine House of Representatives
Education Committee

Dr. Bill Lillard
Superintendent of Schools
Oklahoma City, Oklahoma

Dr. Frederick Mosteller
Chairman, Department of Statistics
Harvard University

Ms. Eleanor P. Sheppard
Chairman, Education Committee
Virginia House of Representatives

Dr. Stephen Wright
Vice President, College Entrance Examination Board

ex officio
Dr. Ralph W. Tyler
Science Research Associates
Chicago, Illinois

APPENDIX D

NATIONAL ASSESSMENT ANALYSIS ADVISORY COMMITTEE

Dr. Frederick Mosteller, Chairman since 1973. Professor of Mathematical Statistics, Harvard University (1970—).

Dr. John W. Tukey, Chairman 1965—73. Professor of Statistics, Princeton University (1965—).

Dr. Robert Abelson, Professor of Psychology, University of North Carolina (1965—72).

Dr. David Brillinger, Professor of Statistics, University of California at Berkeley (1973—).

Dr. William E. Coffman, Professor of Education, University of Iowa (1970—).

Dr. Lee J. Cronbach, Professor of Psychology and Education, Stanford University (1965—69).

Dr. James A. Davis, Director of National Opinion Research Center, Chicago, Illinois (1973—).

Dr. Janet Elashoff, Statistical Advisor, Center for Advanced Study in the Behavioral Sciences, Palo Alto, California (1973—).

Dr. John P. Gilbert, Staff Statistician at the Harvard University Computer Center (1970—).

Dr. Lyle V. Jones, Professor of Psychology, L.L. Thurstone Psychometric Laboratory, Dean of the Graduate School, University of North Carolina (1965—).

Dr. Lincoln Moses, Professor of Statistics, Dean of the Graduate School, Stanford University (1974—).

Dr. Ralph W. Tyler, Science Research Associates, Chicago, Illinois (1965—69).